February 1, 2008
BROWN-HET-1006

FORMATION OF STRUCTURE IN THE UNIVERSE*

Robert H. BRANDENBERGER
Department of Physics
Brown University, Providence, RI 02912, USA
and
Department of Physics
University of British Columbia, Vancouver, BC, V6T 1Z1, CANADA

ABSTRACT

An introduction to modern theories for the origin of structure in the Universe is given. After a brief review of the growth of cosmological perturbations in an expanding Universe and a summary of some important observational results, the lectures focus on the inflationary Universe scenario and on topological defect models of structure formation. A summary of the theory and current observational status of cosmic microwave background temperature fluctuations is given. The final chapter is devoted to some speculative ideas concerning the connection between cosmology and fundamental physics, in particular to ways in which the singularity problem of classical cosmology may be resolved.

* Invited lectures at VIII'th Brazilian School of Cosmology, July 1995; to be published in the proceedings, ed M. Novello (Editions Frontières, Paris 1995).

1. Introduction

Cosmology has over the past fifteen years emerged as a vibrant and exciting subfield of physics. It is based on the marriage of quantum field theory and particle physics on the one hand with classical general relativity on the other. One of the main goals of modern cosmology is to explain the structure of the Universe on the scale of galaxies and beyond. Thus, the experimental/observational basis of the field lies in astronomy, and there is a lot of interaction between theoretical cosmologists and observational astronomers and astrophysicists.

The main goal of these lectures is to give an introduction to the two most developed classes of structure formation theories: those based on inflation and those based on topological defects. I will give a brief survey of relevant observational results from large-scale structure surveys and from searches for cosmic microwave background (CMB) anisotropies, and I will attempt a preliminary comparison with theoretical predictions. A summary of some recent speculative ideas concerning the connection possible solutions of the singularity problem of classical cosmology will be presented. These notes are intended as a pedagogical introduction rather than as a comprehensive review. For comprehensive discussions of inflation, the reader is referred to Refs. 1-3, and for detailed reviews of topological defect models to Refs. 4-7. An introduction to quantum field theory methods used in modern cosmology can be found in Ref. 8. These notes are an updated and expanded version of earlier lecture notes[9] and draw on material presented elsewhere[10] in which some of the topics are treated in more detail.

I hope to persuade the reader that cosmology is an exciting area of physics with close connections to particle and high energy physics, and with a steady stream of new data from astronomy and astrophysics. There is also a close connection with fundamental physics. In fact, cosmology may well be the only arena in which theories such as superstring theory are testable.

The outline of these lectures is as follows: Section 2 is a review of standard cosmology, focusing on its basic principless, its observational support and its prob-

lems.

Section 3 is a brief overview of "new cosmology." I argue why, to obtain an improved cosmological scenario, we need to treat matter using particle physics and field theory. Next, I introduce the basic idea of the inflationary Universe scenario and explain how it leads to a solution of some of the problems of standard cosmology. In particular, it provides a mechanism for the formation of structure in the Universe. The section continues with a brief introduction to the topological defect models of structure formation, an explanation for the need of dark matter, and a survey of the present models.

In Section 4, I present the basics of structure formation, beginning with a survey of some of the relevant large-scale structure data. Structure in the Universe is assumed to grow by gravitational instability. I summarize the essentials of the Newtonian theory of cosmological perturbations (valid on length scales smaller than the apparent horizon (Hubble radius)) and of the relativistic theory[11] (required to study scales beyond the horizon). I also discuss free streaming.

Section 5 contains an overview of inflationary Universe models and of the mechanism for the generation and evolution of perturbations which they provide.

Section 6 presents an overview of topological defect models of structure formation. To begin, a classification of defects is given. Next it is shown that in models which admit topological defects, they are inevitably formed during a symmetry breaking phase transition[12]. Cosmic string and global texture models of structure formation are discussed in detail. In particular, it is pointed out that if defects are responsible for seeding galaxies, there must be new physics at a scale of $\eta \sim 10^{16}$ GeV.

Section 7 focuses on CMB anisotropies. It is shown why theories of structure formation inevitably produce such anisotropies, the predictions of the various models are reviewed, and a comparison with recent observations is given.

Finally, Section 8 contains a summary of a modified theory of gravity in which many of the singularities of classical cosmology can be smoothed out. The theory

is based on a *limiting curvature construction*. Also discussed are some possible connections between superstring theory and cosmology. In particular, a mechanism which might single out three large spatial dimensions is suggested.

In this writeup, units in which $c = \hbar = k_B = 1$ are used unless mentioned otherwise. The space-time metric $g_{\mu\nu}$ is taken to have signature $(+,-,-,-)$. Greek indices run over space and time, latin ones over spatial indices only. The Hubble expansion rate is $H(t) = \dot{a}(t)/a(t)$, with $a(t)$ the scale factor of a Friedmann-Robertson-Walker (FRW) Universe. The present value of H is $100h$ kms^{-1} Mpc^{-1}, where $0.4 < h < 1$. Unless stated otherwise, the value of h is taken to be 0.5. The cosmological redshift at time t is denoted by $z(t)$. As usual, the symbols G and m_{pl} stand for Newton's constant and Planck mass, respectively. Distances are measured in pc ("parsec") or Mpc, where 1 pc corresponds to 3.1 light years.

2. Review of Standard Cosmology

2.1. Principles

The standard big bang cosmology rests on three theoretical pillars: the cosmological principle, Einstein's general theory of relativity and a perfect fluid description of matter.

The cosmological principle[13] states that on large distance scales the Universe is homogeneous. From an observational point of view this is an extremely nontrivial statement. On small scales the Universe looks rather inhomogeneous. The inhomogeneities of the solar system are obvious to everyone, and even by the naked eye it is apparent that stars are not randomly distributed. They are bound into galaxies, dynamical entities whose visible radius is about 10^4 pc. Telescopic observations show that galaxies are not randomly distributed, either. Dense clumps of galaxies can be identified as Abell clusters. In turn, Abell cluster positions are correlated to produce the large-scale structure dominated by sheets (or filaments), with typical scale 100 Mpc, observed in recent redshift surveys[14]. Until recently,

every new survey probing the Universe to greater depth revealed new structures on the scale of the sample volume. In terms of the visible distribution of matter there was no evidence for large-scale homogeneity. This situation changed in 1992 with the announcement[15] that a new redshift survey, complete to a depth of about $500h^{-1}$ Mpc, had discovered no prominent structures on scales larger than $100h^{-1}$ Mpc. This is the first observational evidence from optical measurements in favor of the cosmological principle. However, to put this result in perspective we must keep in mind that the observed isotropy of the CMB temperature[16] to better than 10^{-5} on large angular scales has been excellent evidence for the validity of the cosmological principle.

The second theoretical pillar is general relativity, the theory which determines the dynamics of the Universe. According to the cosmological principle, space at any time t is a three dimensional surface with maximal symmetry (translations and rotations). There are three families of such spaces[17]: flat Euclidean space R^3, the three sphere S^3, and the hypersphere H^3. The proper distance ds^2 on these three surfaces can be written in spherical coordinates as

$$ds^2 = a(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2(d\vartheta^2 + \sin^2\vartheta d\varphi^2)\right] . \tag{2.1}$$

The constant k is $+1$, 0, or -1 respectively for S^3, R^3 and H^3.

The Einstein equations of general relativity imply that $a(t)$ – called the scale factor of the Universe – evolves in time. The proper distance/time in space-time is

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2(d\vartheta^2 + \sin^2\vartheta d\varphi^2)\right] . \tag{2.2}$$

By a coordinate transformation, $a(t)$ can be set equal to 1 at the present time t_0.

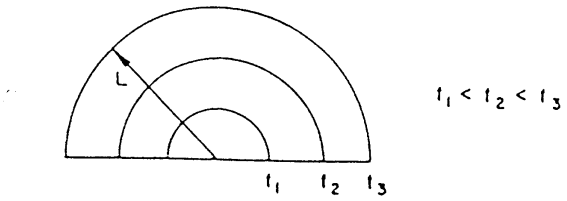

Figure 1: Sketch of the expanding Universe. Concentric circles indicate space at fixed time, with time increasing as the radius gets larger. Points at rest have constant comoving coordinates. Their world lines are straight lines through the origin (e.g. L).

To obtain a simple visualization of an expanding Universe, consider space to be the surface of a balloon. We draw a grid on the surface and use it to define coordinates $\underline{x}^c$ (the superscript c stands for comoving). Points at rest on the surface of the balloon have constant comoving coordinates. However, if the balloon is being inflated, then the physical distance Δx^p betwen two points at rest with comoving separation Δx^c increases:

$$\Delta x^p = a(t)\Delta x^c \,. \tag{2.3}$$

The scale factor $a(t)$ is proportional to the radius of the balloon (see Fig. 1).

According to Einstein's equivalence principle, particles in the absence of external nongravitational forces move on geodesies, curves which extremize ds^2. The velocity of a particle relative to the expansion of the Universe is called peculiar velocity v_p

$$v_p = a(t)\,\frac{dx^c}{dt} \tag{2.4}$$

and obeys the equation

$$\ddot{v}_p + \frac{\dot{a}}{a}\,v_p = 0\,, \tag{2.5}$$

from which it follows that

$$v_p(t) \sim a^{-1}(t)\,. \tag{2.6}$$

The dynamics of an expanding Universe is determined by the Einstein equa-

tions, which relate the expansion rate to the matter content, specifically to the energy density ρ and pressure p. For a homogeneous and isotropic Universe, they reduce to the Friedmann-Robertston-Walker (FRW) equations

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{k}{a^2} = \frac{8\pi G}{3}\rho \tag{2.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p\right). \tag{2.8}$$

These equations can be combined to yield the continuity equation (with Hubble constant $H = \dot{a}/a$)

$$\dot{\rho} = -3H(\rho + p)\,. \tag{2.9}$$

The third key assumption of standard cosmology is that matter is described by an ideal gas with an equation of state

$$p = w\rho\,. \tag{2.10}$$

For cold matter, pressure is negligible and hence $w = 0$. From (2.9) it follows that

$$\rho_m(t) \sim a^{-3}(t)\,, \tag{2.11}$$

where ρ_m is the energy density in cold matter. For radiation we have $w = 1/3$ and hence it follows from (2.9) that

$$\rho_r(t) \sim a^{-4}(t)\,, \tag{2.12}$$

$\rho_r(t)$ being the energy density in radiation.

2.2. Observational Pillars

The first observational pillar of standard cosmology is Hubble's redshift-distance relationship[18] (Fig. 2)

$$z = Hd\,, \tag{2.13}$$

where H is the present Hubble expansion constant, d is the distance to a galaxy, and z is its redshift

$$z \equiv \frac{\lambda_0}{\lambda_e} - 1\,, \tag{2.14}$$

$\lambda_e(\lambda_0)$ being the wavelength of light at the time of emission (detection).

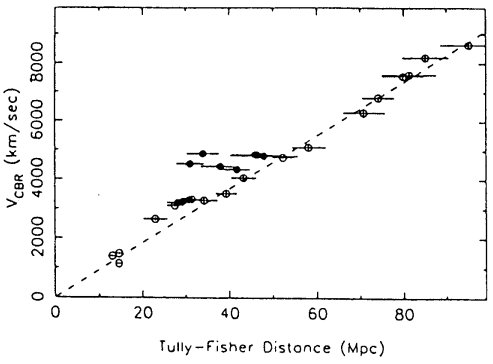

Figure 2: A recent redshift-distance plot of galaxies[19]. The distances are determined using the Tully-Fisher method. See Ref. 31 for a detailed discussion of the method and errors.

There is an easy intuitive derivation of this result. A wave in an expanding background will have a wavelength which increases as the scale factor $a(t)$. Hence for light emitted at time t_e

$$z(t_e) = \frac{a(t_0)}{a(t_e)} - 1\,. \tag{2.15}$$

For light emitted close to the present time we can Taylor expand the above result

to obtain

$$z(t_e) \simeq \frac{\dot{a}(t_0)}{a(t_0)} (t_0 - t_e) \simeq H(t_0)\, d\,. \tag{2.16}$$

Equation (2.15) defines the cosmological redshift, which can be used as a measure of cosmic time.

The second observational pillar of standard cosmology is the existence and black body nature of the CMB[20,21)]

To understand the connection[17)], consider matter in an expanding Universe. As we go backwards in time, the density of matter increases as $a^{-3}(t)$, and as a consequence the temperature grows. Above a temperature of 13.6 eV, atoms are ionized, and a bath of photons in thermal equilibrium must be present. Photons still scatter frequently below this temperature. At some time t_{rec} the scattering length of a photon becomes longer than the Hubble radius. After that, photons travel without scattering. At t_{rec}, the distribution of photons is of black body type. A special feature of black body spectra is that the spectral shape is maintained even after t_{rec}. The only change is that the temperature redshifts

$$T(t) = \frac{a(t_{rec})}{a(t)}\, T(t_{rec})\,. \tag{2.17}$$

Hence, the standard Big Bang model predicts a black body spectrum of photons with temperature

$$T_o = T_{rec} z(t_{rec})^{-1}\,, \tag{2.18}$$

where T_{rec} is the temperature at t_{rec}, determined by comparing the largest rate of scattering of photons below recombination, that due to Thomson scattering, with the Hubble expansion rate, yielding the result

$$T_{rec} \simeq 0.25\,\mathrm{eV} \simeq 4000^{\circ}\mathrm{K}\,. \tag{2.19}$$

The corresponding redshift is determined by measuring T_o.

In 1965, Penzias and Wilson[22] discovered this remnant black body radiation at a temperature of about 3° K. Since the spectrum peaks in the microwave region it is now called the CMB (cosmic microwave background). Recent satellite (COBE)[23] and rocket[24] experiments have confirmed the black body nature of the CMB to very high accuracy. The temperature is 2.73° K$= T_0$ which corresponds to

$$z(t_{rec}) = z_{rec} \sim 10^3 \,. \tag{2.20}$$

Given the existence of the CMB, we know that matter has two components: dust (with energy density $\rho_m(t)$) and radiation (with density $\rho_r(t)$). At the present time t_0, $\rho_m(t) \gg \rho_r(t)$. The radiation energy density is determined by T_0, and the matter energy density can be estimated by analyzing the dynamics of galaxies and clusters and using the virial theorem. However, since by (2.11) and (2.12) $\rho_m(t) \sim a(t)^{-3}$ and $\rho_r(t) \sim a(t)^{-4}$, as we go back in time the fraction of energy density in radiation increases, and the two components become equal at a time t_{eq}, the time of equal matter and radiation. The corresponding redshift is

$$z_{eq} \simeq \Omega \, h_{50}^{-2} \, 10^4 \tag{2.21}$$

where

$$\Omega = \frac{\rho}{\rho_c} \, (t_0) \,, \tag{2.22}$$

ρ_c being the density for a spatially flat Universe (the critical density), and h_{50} is the value of H in units of 50 km s^{-1} Mpc^{-1}.

The time t_{eq} is important for structure formation. As we will see in Section 4, it is only after t_{eq} that perturbations on scales smaller than the Hubble radius $H^{-1}(t)$ can grow. Before then, the radiation pressure prevents growth. A temperature-time plot of the early Universe is sketched in Fig. 3, Note that $t_{eq} < t_{rec}$.

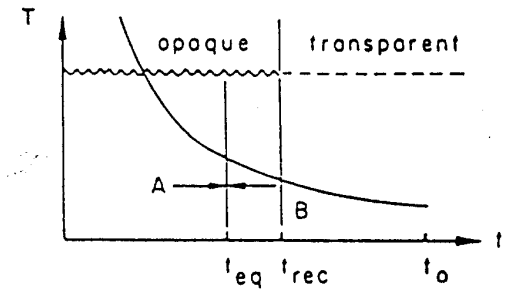

Figure 3: Temperature-time diagram of standard big bang cosmology. The present time, time of last scattering and time of equal matter and radiation are t_0, t_{rec} and t_{eq} respectively. The Universe is radiation-dominated before t_{eq} (Region A) and matter-dominated in Region B. Before and after t_{rec}, respectively, the Universe was opaque and transparent, respectively, to microwave photons.

The third observational pillar of standard big bang cosmology concerns nucleosynthesis[25,26] - the production of light elements (heavy elements are formed in supernovae). Above a temperature of about 10^9 K, the nuclear interactions are sufficiently fast to prevent neutrons and protons from fusing. However, below that temperature, it is thermodynamically favorable for neutrons and protrons to fuse and form deuterium, helium 3, helium 4 and lithium 7 through a long and interconnected chain of reactions. The resulting light element abundances depend sensitively on the expansion rate of the Universe and on Ω_B, the fraction of energy density ρ_B at present in baryons relative to the critical density ρ_c. In Fig. 5, recent theoretical calculations[27] of the abundances are shown and compared with observations. Demanding agreement with all abundances leaves only a narrow window

$$3 \times 10^{-10} < \eta < 10^{-9}\,, \tag{2.23}$$

where η is the ratio of baryon number density n_B to photon number density n_γ

$$\eta = \frac{n_b}{n_\gamma}\,. \tag{2.24}$$

From (2.23), it follows that Ω_B is constrained:

$$0.01 < \Omega_B h^2 < 0.035\,. \tag{2.25}$$

In particular, if the Universe is spatially flat and the cosmological constant is negligible, there must be nonbaryonic dark matter. We will return to the dark matter issue in Section 3.

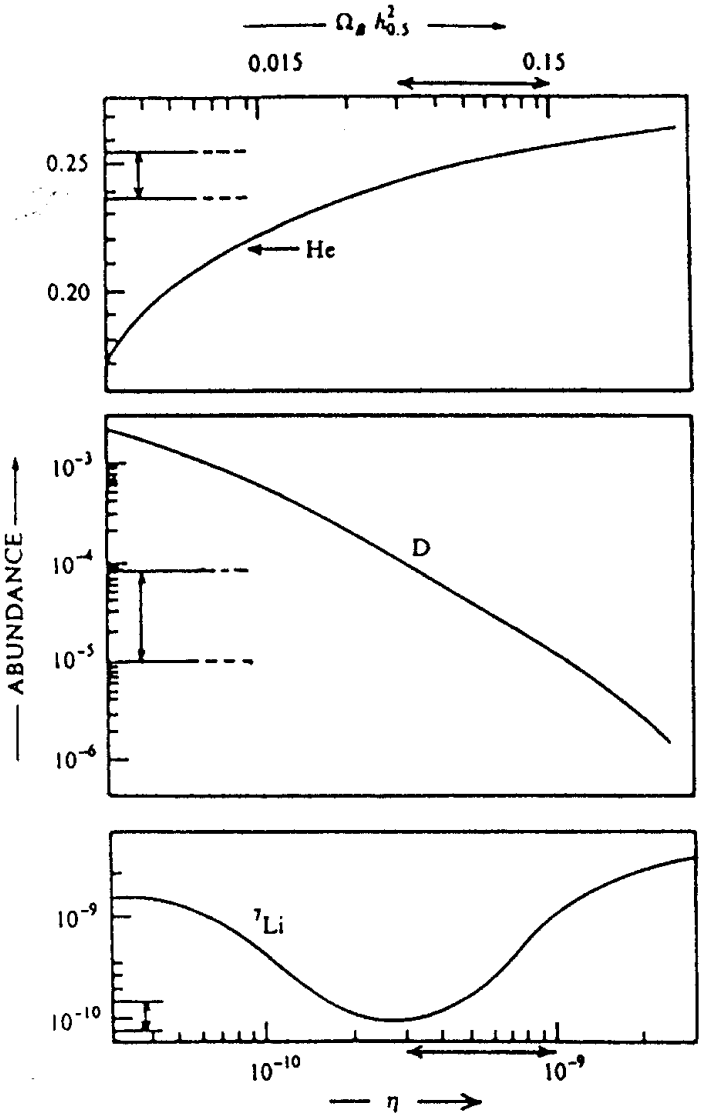

Figure 4: Light el-

ement abundances as a function of the baryon to entropy ratio η (from Ref. 27). The solid curves are the predictions of homogeneous big bang nucleosynthesis. The observational limits are indicated on the left vertical axis. Theory and observations are only consistent for a narrow range of values of η.

The final pillar of standard cosmology is the near isotropy of the CMB[16]. After subtracting the dipole anisotropy which is presumed to be due to the motion of the earth relative to the rest frame defined by the CMB, no anisotropies have been detected to a level of better than 10^{-4}, i.e., the temperature difference $\delta T(\vartheta)$ between two beams pointing in directions in the sky separated by an angle ϑ (Fig. 4) satisfies

$$\frac{\delta T(\vartheta)}{\bar{T}} < 10^{-4} \tag{2.26}$$

on all angular scales ϑ. Here $\bar{T}$ is the average temperature.

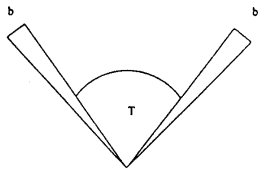

Figure 5: Sketch of a CMB anisotropy experiment. Two radio antennas with beam width b collect microwave radiation from points in the sky separated by an angle T. The difference in beam intensities is measured.

Until recently, the isotropy of the CMB was the only observational support for the cosmological principle. Any inhomogeneities of the Universe on length scales comparable to the comoving Hubble radius at t_{rec} and larger would generate temperature anisotropies by a mechanism discussed in detail in Section 7. Hence, the near isotropy of the CMB implies that density fluctuations on large scales must have been very small in the early Universe.

To summarize, the observational pillars of standard cosmology are Hubble's redshift-distance relation, the existence and black body nature of the CMB, primordial nucleosynthesis, and the isotropy of the CMB. Note, in particular, that

no tests of big bang cosmology say anything about the evolution of the Universe before the time of nucleosynthesis. Note, also, that not all astronomers accept the above observations as support of the Big Bang model. For a recent criticism see Ref. 28 (and Ref. 29 for a reply to the criticism).

2.3. Problems

Standard Big Bang cosmology is faced with several important problems. Only one of these, the age problem, is a potential conflict with observations. The others which I will focus on here – the homogeneity, flatness and formation of structure problems (see e.g. Ref. 30) – are questions which have no answer within the theory and are therefore the main motivation for the new cosmological models which will be discussed in the rest of these lectures.

From the FRW equations (2.7) and (2.8) it is easy to calculate the age of the Universe, given the expansion law (2.11) which holds throughout most of the history of the Universe. For a spatially flat Universe, the age τ depends on the expansion rate H, i.e. on the constant h which is in the range $0.4 < h < 1$:

$$\tau \simeq \frac{7}{h} 10^9 \mathrm{yr} \,. \tag{2.27}$$

Globular cluster ages have been estimated to lie in the range $12 - 18 \times 10^9$ yr. Thus, theory and observations are only consistent if $h < 0.55$ (see e.g. Ref. 31). In an open Universe the problem is less severe. Recent observations have not led to a decrease in the uncertainty in the value of h. Observations by the Hubble space telescope[209] and on supernovae observations[210] indicate a fairly large value ($h \simeq 0.8$), but direct measurements based on the Sunyaev-Zeldovich effect and using more distant galaxy clusters yield[211] a small value ($h \simeq 0.5$). Modern cosmological models do not add any insight into the age problem since they only modify the evolution of the Universe at very early times $t \ll t_{eq}$.

The final three problems mentioned above, the homogeneity, flatness and formation of structure problems, provided a lot of the motivation for the development of the inflationary Universe scenario[30] and will hence be discussed in detail.

The horizon problem is illustrated in Fig. 6. As is sketched, the comoving region $\ell_p(t_{rec})$ over which the CMB is observed to be homogeneous to better than one part in 10^4 is much larger than the comoving forward light cone $\ell_f(t_{rec})$ at t_{rec}, which is the maximal distance over which microphysical forces could have caused the homogeneity:

$$\ell_p(t_{rec}) = \int_{t_{rec}}^{t_0} dt\, a^{-1}(t) \simeq 3\, t_0 \left(1 - \left(\frac{t_{rec}}{t_0}\right)^{1/3}\right) \tag{2.28}$$

$$\ell_f(t_{rec}) \int_0^{t_{rec}} dt\, a^{-1}(t) \simeq 3\, t_0^{2/3}\, t_{rec}^{1/3}\,. \tag{2.29}$$

From the above equations it is obvious that $\ell_p(t_{rec}) \gg \ell_f(t_{rec})$. Hence, standard cosmology cannot explain the observed isotropy of the CMB.

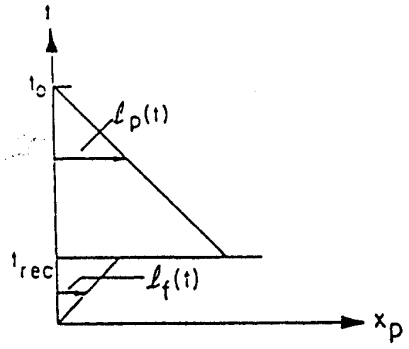

Figure 6: A space-time diagram (physical distance x_p versus time t) illustrating the homogeneity problem: the past light cone $\ell_p(t)$ at the time t_{rec} of last scattering is much larger than the forward light cone $\ell_f(t)$ at t_{rec}.

In standard cosmology and in an expanding Universe, $\Omega = 1$ is an unstable fixed point. This can be seen as follows. For a spatially flat Universe ($\Omega = 1$)

$$H^2 = \frac{8\pi G}{3}\,\rho_c\,, \tag{2.30}$$

whereas for a nonflat Universe

$$H^2 + \varepsilon\, T^2 = \frac{8\pi G}{3}\,\rho\,, \tag{2.31}$$

with

$$\varepsilon = \frac{k}{(aT)^2}\,. \tag{2.32}$$

The quantity ε is proportional to $s^{-2/3}$, where s is the entropy density. Hence, in standard cosmology, ε is constant. Combining (2.30) and (2.31) gives

$$\frac{\rho - \rho_c}{\rho_c} = \frac{3}{8\pi G}\,\frac{\varepsilon T^2}{\rho_c} \sim T^{-2}\,. \tag{2.33}$$

Thus, as the temperature decreases, $\Omega - 1$ increases. In fact, in order to explain the present small value of $\Omega - 1 \sim \mathcal{O}(1)$, the initial energy density had to be extremely close to critical density. For example, at $T = 10^{15}$ GeV, (2.33) implies

$$\frac{\rho - \rho_c}{\rho_c} \sim 10^{-50}\,. \tag{2.34}$$

What is the origin of these fine tuned initial conditions? This is the flatness problem of standard cosmology.

The last problem of the standard cosmological model I will mention is the "formation of structure problem." Observations indicate that galaxies and even clusters of galaxies have nonrandom correlations on scales larger than 50 Mpc (see e.g. Ref. 14). This scale is comparable to the comoving horizon at t_{eq}. Thus, if the initial density perturbations were produced much before t_{eq}, the correlations cannot be explained by a causal mechanism. Gravity alone is, in general, too weak to build up correlations on the scale of clusters after t_{eq} (see, however, the explosion scenario of Ref. 32). Hence, the two questions of what generates the primordial density perturbations and what causes the observed correlations, do not have an answer in the context of standard cosmology. This problem is illustrated by Fig. 7.

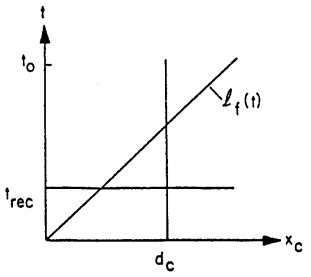

Figure 7: A sketch (conformal separation vs. time) of the formation of structure problem: the comoving separation d_c between two clusters is larger than the forward light cone at time t_{eq}.

Finally, let us address the cosmological constant problem. All known symmetries of nature and principles of general relativity allow for the presence of a term in the Einstein equations which acts like matter with energy Λ and pressure $-\Lambda$, i.e., with an equation of state $p = -\rho$. If it is not to dominate the present expansion rate of the Universe, the cosmological constant Λ must be very small

$$\Lambda < 3H_0^2 \sim 10^{-83}\mathrm{GeV}^2 \,. \tag{2.35}$$

On dimensional grounds, we would expect Λ to be of the order $m_{pl}^2 \sim 10^{38}\,\mathrm{GeV}^2$. Thus, the cosmological constant is about 140 orders of magnitude smaller than what we would expect it to be (for recent reviews of the cosmological constant problem, see Ref. 33).

As we will see, modern cosmology does not address the cosmological constant problem. If anything, the problem will manifest itself in a more apparent manner. For some recent ideas on how infrared effects in field theory might solve the cosmological constant problem see Ref. 212.

Due to the formation of structure problem, there can be no causal physical theory for the origin of structure (with nontrivial spatial correlations) in the Universe in the context of the Standard Big Bang theory. The main breakthrough of modern cosmology is that it provides solutions to this problem. The key to understanding this breakthrough in cosmology is the realization of the internal inconsistency

of the standard picture when extrapolated to times much before nucleosynthesis. Standard cosmology is based on the assumption that matter continues to be described by an ideal radiation gas to arbitrarily high temperatures. This is clearly in contrast to what nuclear and particle physics tells us. As we go backwards in time towards the Big Bang, nuclear physics and eventually particle physics effects will take over. To describe matter correctly, a quantum field theoretic description must be used. Note, however, that at a fundamental level there is an inconsistency if matter is described quantum mechanically while maintaining a classical description of gravity. Hence, we cannot hope that any of the present cosmological theories will be the ultimate theory.

3. New Cosmology and Structure Formation

The goal of this section is to present an overview of what can be gained if we go beyond standard cosmology and allow matter to be described in terms of concepts from particle physics. Detailed discussions of the models will be given in later sections.

3.1. The Inflationary Unvierse

The idea of inflation[30] is very simple. We assume there is a time interval beginning at t_i and ending at t_R (the "reheating time") during which the Universe is exponentially expanding, i.e.,

$$a(t) \sim e^{Ht}, \quad t\epsilon\,[t_i,\, t_R] \tag{3.1}$$

with constant Hubble expansion parameter H. Such a period is called "de Sitter" or "inflationary." The success of Big Bang nucleosynthesis sets an upper limit to the time of reheating:

$$t_R \ll t_{NS}\,, \tag{3.2}$$

t_{NS} being the time of nucleosynthesis.

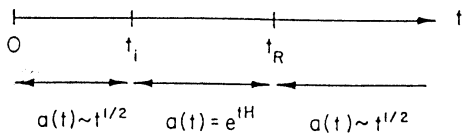

Figure 8: The phases of an inflationary Universe. The times t_i and t_R denote the beginning and end of inflation, respectively. In some models of inflation, there is no initial radiation domintated FRW period. Rather, the classical space-time emerges directly in an inflationary state from some initial quantum gravity state.

The phases of an inflationary Universe are sketched in Fig. 8. Before the onset of inflation there are no constraints on the state of the Universe. In some models a classical space-time emerges immediately in an inflationary state, in others there is an initial radiation dominated FRW period. Our sketch applies to the second case. After t_R, the Universe is very hot and dense, and the subsequent evolution is as in standard cosmology. During the inflationary phase, the number density of any particles initially in thermal equilibrium at $t = t_i$ decays exponentially. Hence, the matter temperature $T_m(t)$ also decays exponentially. At $t = t_R$, all of the energy which is responsible for inflation (see later) is released as thermal energy. This is a nonadiabatic process during which the entropy increases by a large factor. The temperature-time evolution in an inflationary Universe is depicted in Fig. 9.

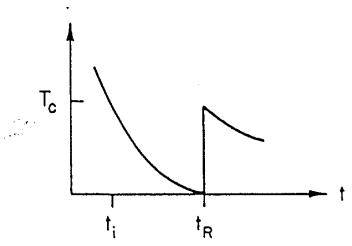

Figure 9: The time dependence of matter temperature in an inflationary Universe. During the period of exponential expansion, the temperature decreases exponentially. At the end of inflation the energy density of the scalar field responsible for inflation is transferred to ordinary matter. This leads to reheating. The critical temperature T_c is the temperature at which

the initial matter thermal energy density becomes less than the scalar field energy density (see Chapter 5).

Fig. 10 is a sketch of how a period of inflation can solve the homogeneity problem. $\Delta t = t_R - t_i$ is the period of inflation. During inflation, the forward light cone increases exponentially compared to a model without inflation, whereas the past light cone is not affected for $t \geq t_R$. Hence, provided Δt is sufficiently large, $\ell_f(t_R)$ will be greater than $\ell_p(t_R)$. The condition on Δt depends on the temperature T_R corresponding to time t_R, the temperature of reheating. Demanding that $\ell_f(t_R) > \ell_p(t_R)$ we find, using the analogs of (2.28) and (2.29), the following criterion

$$e^{\Delta t H} \geq \frac{\ell_p(t_R)}{\ell_f(t_R)} \simeq \left(\frac{t_0}{t_R}\right)^{1/2} = \left(\frac{T_R}{T_0}\right) \sim 10^{27} \tag{3.3}$$

for $T_R \sim 10^{14}$GeV and $T_0 \sim 10^{-13}$GeV (the present microwave background temperature). Thus, in order to solve the homogeneity problem, a period of inflation with

$$\Delta t \gg 50\, H^{-1} \tag{3.4}$$

is required.

Figure 10: Sketch (physical coordinates vs. time) of the solution of the homogeneity problem. During inflation, the forward light cone $l_f(t)$ is expanded exponentially when measured in physical coordinates. Hence, it does not require many e-foldings of inflation in order that $l_f(t)$ becomes larger than the past light cone at the time of last scattering. The dashed line is the forward light cone without inflation.

Inflation also can solve the flatness problem[34,30] The key point is that the entropy density s is no longer constant. As will be explained later, the temperatures at t_i and t_R are essentially equal. Hence, the entropy increases during inflation

by a factor $\exp(3H\Delta t)$. Thus, ϵ decreases by a factor of $\exp(-2H\Delta t)$. With the numbers used in (3.3):

$$\epsilon_{\rm after} \sim 10^{-54}\, \epsilon_{\rm before}\,. \tag{3.5}$$

Hence, $(\rho - \rho_c)/\rho$ can be of order 1 both at t_i and at the present time. In fact, if inflation occurs at all, the theory then predicts that at the present time $\Omega = 1$ to a high accuracy (now $\Omega < 1$ would require special initial conditions).

What was said above can be rephrased geometrically: during inflation, the curvature radius of the Universe – measured on a fixed physical scale – increases exponentially. Thus, a piece of space looks essentially flat after inflation even if it had measurable curvature before.

Most importantly, inflation provides a mechanism which in a casual way generates the primordial perturbations required for galaxies, clusters and even larger objects. In inflationary Universe models, the Hubble radius ("apparent" horizon), $3t$, and the "actual" horizon (the forward light cone) do not coincide at late times. Provided (3.3) is satisfied, then (as sketched in Fig. 11) all scales within our apparent horizon were inside the actual horizon since t_i. Thus, it is in principle possible to have a casual generation mechanism for perturbations[35–38].

Figure 11: A sketch (physical coordinates vs. time of the solution of the formation of structure problem. Provided that the period of inflation is sufficiently long, the separation d_c between two galaxy clusters is at all times smaller than the forward light cone. The dashed line indicates the Hubble radius. Note that d_c starts out smaller than the Hubble radius, crosses it during the de Sitter period, and then reenters it at late times.

The generation of perturbations is supposed to be due to a causal microphysical process. Such processes can only act coherently on length scales smaller than the Hubble radius $\ell_H(t)$ where

$$\ell_H(t) = H^{-1}(t)\,. \tag{3.6}$$

A heuristic way to understand the meaning of $\ell_H(t)$ is to realize that it is the distance which light (and hence the maximal distance any causal effects) can propagate in one expansion time:

$$\ell_H(t) \sim a(t) \int_t^{t+H^{-1}(t)} a(t')^{-1}\, dt' \,. \tag{3.7}$$

In Section 5 a more mathematical justification for the definition and role of $\ell_H(t)$ will be given.

As will be discussed in Section 5, the density perturbations produced during inflation are due to quantum fluctuations in the matter and gravitational fields[36,37]. The amplitude of these inhomogeneities corresponds to a tempertuare T_H

$$T_H \sim H \,, \tag{3.8}$$

the Hawking temperature of the de Sitter phase. This implies that at all times t during inflation, perturbations with a fixed physical wavelength $\sim H^{-1}$ will be produced. Subsequently, the length of the waves is streched with the expansion of space, and soon becomes larger than the Hubble radius. The phases of the inhomogeneities are random. Thus, the inflationary Universe scenario predicts perturbations on all scales ranging from the comoving Hubble radius at the beginning of inflation to the corresponding quantity at the time of reheating. In particular, provided that inflation lasts sufficiently long (see (3.4)), perturbations on scales of galaxies and beyond will be generated. Note, however, that it is very dangerous to interpret de Sitter Hawking radiation as thermal radiation. In fact, the euation of state of this "radiation" is not thermal[213].

Now that the reader is (hopefully) convinced that inflation is a beautiful idea, the question arises how to realize this scenario. The initial hope was that the same scalar fields (Higgs fields) which particle physicists introduce in order to spontaneously break the internal symmetries of their field theory models would

lead to inflation. This hope was based on the fact that the energy density ρ and pressure p of a real scalar field $\varphi(\underline{x}, t)$ are given by

$$\begin{aligned} \rho(\varphi) &= \frac{1}{2}\,\dot{\varphi}^2 + \frac{1}{2}\,(\nabla\varphi)^2 + V(\varphi) \\ p(\varphi) &= \frac{1}{2}\dot{\varphi}^2 - \frac{1}{6}(\nabla\varphi)^2 - V(\varphi)\,. \end{aligned} \tag{3.9}$$

Thus, provided that at some initial time t_i

$$\dot{\varphi}(\underline{x},\, t_i) = \nabla\varphi(\underline{x}_i\, t_i) = 0 \tag{3.10}$$

and

$$V(\varphi(\underline{x}_i\, t_i)) > 0\,, \tag{3.11}$$

the equation of state of matter will read

$$p = -\rho \tag{3.12}$$

and, from the FRW equations it will follow that

$$a(t) = e^{tH}\,,\; H^2 = \frac{8\pi G}{3}\,V(\varphi)\,. \tag{3.13}$$

The next question is how to realize the required initial conditions (3.10) and to maintain the key constraints

$$\dot{\varphi}^2 \ll V(\varphi)\,,\; (\nabla\varphi)^2 \ll V(\varphi) \tag{3.14}$$

for sufficiently long (see (3.4)). This typically requires potentials which are very flat near $\varphi(\underline{x},\, t_i)$. Worse yet, the parameters of the potential $V(\varphi)$ must be chosen such that the final amplitude of density perturbations is sufficiently small to agree with the limits on CMB anisotropies. As we will demonstrate in Section 5, these conditions impose severe constraints on the constants which appear in $V(\varphi)$.

In light of these difficulties it is important to keep in mind that inflation can also be generated by modifying gravity at high curvatures (see e.g., Refs. 39-41). It is also wise to investigate alternative theories of structure formation which do not rely on inflation.

3.2. Topological Defect Models

According to particle physics theories, matter at high energies and temperatures must be described in terms of fields. Gauge symmetries have proved to be extremely useful in describing the standard model of particle physics, according to which at high energies the laws of nature are invariant under a nonabelian group G of internal symmetry transformations

$$G = \mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times U(1)_Y \tag{3.15}$$

which at a temperature of about 200 MeV is spontaneously broken down to

$$G' = \mathrm{SU}(3)_c \times \mathrm{U}(1)\,. \tag{3.16}$$

The subscript on the SU(3) subgroup indicates that it is the color symmetry group of the strong interactions, $\mathrm{SU}(2)_L\times\ \mathrm{U}(1)_Y$ is the Glashow-Weinberg-Salam (WS) model of weak and electromagnetic interactions, the subscripts L and Y denoting left handedness and hypercharge respectively. At low energies, the WS model spontaneously breaks to the U(1) subgroup of electromagnetism.

Spontaneous symmetry breaking is induced by an order parameter φ taking on a nontrivial expectation value $<\varphi>$ below a certain temperture T_c. In some particle physics models, φ is a fundamental scalar field in a nontrivial representation of the gauge group G which is broken. However, φ could also be a fermion condensate, as in the BCS theory of superconductivity.

The transition taking place at $T = T_c$ is a phase transition and T_c is called the critical temperature. From condensed matter physics it is well known that in many cases topological defects form during phase transitions, particularly if the transition rate is fast on a scale compared to the system size. When cooling a metal, defects in the crystal configuration will be frozen in; during a temperature quench of ^{4}He, thin vortex tubes of the normal phase are trapped in the superfluid; and

analogously in a temperature quench of a superconductor, flux lines are trapped in a surrounding sea of the superconducting Meissner phase (see Fig. 12 for an example[42] of defect formation).

Figure 12: A simulation of defect formation in the $2+1$ dimensional Abelian Higgs model[42] in an expanding background. The total energy density is plotted against the two spatial coordinates. The initial conditions were specified by thermal initial conditions with random phases of the order parameter on Hubble scales. At later times, the thermal noise has redshifted away, leaving behind trapped energy density in vortices.

In cosmology, the rate at which the phase transition proceeds is given by the expansion rate of the Universe. Hence, topological defects will inevitably be produced in a cosmological phase transition[12], provided the underlying particle physics model allows such defects.

Topological defects can be point-like (monopoles), string-like (cosmic strings)[43] or planar (domain walls), depending on the particle physics model (see Section 6). Also of importance are textures[44,45], point defects in space-time.

Topological defects represent regions in space with trapped energy density. These regions of surplus energy can act as seeds for structure formation as is illustrated in Fig. 13. For point-like defects, the force which causes clustering about the seed can be understood using Newtonian gravity. The process is called gravitational accretion. For precise calculations (and in the case of other defects), general relativistic effects must be taken into account (see Section 6).

Figure 13: Sketch of the basic gravitational accretion mechanism. The topological defect (in this case a cosmic string loop) is a configuration of trapped energy density. This excess density produces a Newtonian gravitational attractive force on the surrounding matter.

No stable topological defects arise in the breaking of the WS model. However, there is good evidence for phase transitions at very high energies. The coupling

constants of $SU(3)_c$, SU(2) and U(1) are seen to converge at an energy scale η of about

$$\eta \sim 10^{16}\,\mathrm{GeV}\,. \tag{3.17}$$

It is therefore not unreasonable to speculate that the standard model results from the breaking of a larger symmetry group G_0 at a scale η. A large class of unified gauge theories[46] based on a symmetry breaking

$$G_0 \longrightarrow G = \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \tag{3.18}$$

admit topological defects, many theories have cosmic string solutions.

Topological defect models of structure formation will be discussed in detail in Section 6. Here I will briefly point out by which mechanism correlations on all cosmological scales are induced. To be concrete, I consider the cosmic string model[47,48]. Cosmic strings are one-dimensional defects without ends. Hence, they must be either infinite in length or else closed loops. The fact that at the time of the phase transition t_c, the order parameter has random phases on scales larger than the initial correlation length implies that at t_c a random walk-like network of infinite strings will form[12]. This implies nontrivial correlations of structures seeded by these strings on all scales larger than the initial correlation length.

3.3. Need for Dark Matter

At this point we have illustrated two classes of mechanisms by which structure formation in the Universe can be seeded: quantum fluctuations during a period of inflation, and topological defects. To completely specify a theory of structure formation, however, we must also specify what the "dark matter" which dominates the energy density of the Universe today consists of (see e.g. Ref. 49).

The evidence for dark matter has been accumulating over the past decade. Measurements of galaxy velocity rotation curves indicate that a large fraction of

the mass of a galaxy does not shine. In Fig. 14, the measured rotation velocity v is plotted as a function of the distance r from the center of the galaxy. This velocity is compared to the velocity $\hat{v}(r)$ which results from the virial theorem, assuming that light traces mass (the luminosity profile is given in the upper frame). The data is for the spiral galaxies NGC2403 and NGC3198[50]. The comparison shows that the mass of these galaxies extends significantly beyond the visible radius and that – assuming Newtonian gravity is applicable – a large fraction of the mass of a spiral galaxy must be dark.

Figure 14: Velocity rotation curves for two galaxies (from Ref. 50). The upper panel shows the luminosity of the galaxy as a function of the distance from the center, the lower panel presents the velocity (vertical axis, in units of kms^{-1}) as a function of the same distance. The solid curve is the velocity inferred from the observed luminosity curve, using the virial theorem, the dotted curves are the observational results. The fact that the velocity rotation curves remain constant beyond the visible radius of the galaxy is strong evidence for galactic dark matter.

The fraction of matter that shines can be expressed as a fraction $\Omega_{\rm lum}$ of the critical density ρ_c. The present estimates give

$$\Omega_{\rm lum} \ll 0.01\,, \tag{3.19}$$

whereas the fraction Ω_g of mass in galaxies is much larger[51]

$$\Omega_g \sim 0.03\,. \tag{3.20}$$

It is also possible to estimate the fraction Ω_{cl} of mass which is gravitationally bound in clusters. Current estimates using the virial theorem give[51]

$$\Omega_{cl} \sim 0.1 - 0.2 \tag{3.21}$$

(this comes mainly from studying the infall of galaxies towards the Virgo cluster).

Finally, the amount Ω_{LSS} of mass in large-scale structures can be estimated by measuring the large-scale peculiar velocities and inferring the mass required to generate such velocities. Current estimates give[52,53]

$$\Omega_{LSS} = 0.8 \pm 0.5\,. \tag{3.22}$$

As mentioned in Section 2, nucleosynthesis provides independent limits on the fraction Ω_B of mass in baryons (see (2.25)). Comparing (3.19-3.22) with (2.25) we conclude:

i) There must be baryonic dark matter. In fact, most of the dark matter in galaxies and/or clusters could be baryonic, and some of this baryonic dark matter may have recently been discovered by gravitational microlensing.

ii) If $\Omega = 1$, then most of the matter in the Universe consists of nonbaryonic dark matter. In fact, there is increasing evidence (see (3.22)) that nonbaryonic dark matter must exist independent of the theoretical prejudice for $\Omega = 1$.

The dark matter in the Universe is visible only through its gravitational effects. Hence, nonbaryonic dark matter candidates can be divided into two classes, cold dark matter (CDM) and hot dark matter (HDM).

CDM particles are cold, i.e., their peculiar velocity v is negligible at the time t_{eq} when structure formation begins:

$$v(t_{eq}) \ll 1\,. \tag{3.23}$$

Candidates for CDM include the axion (coherent oscillations of a low mass scalar field) and neutralinos (the lightest stable supersymmetric particle, which must be neutral).

HDM particles are relativistic at t_{eq}:

$$v(t_{eq}) \sim 1\,. \tag{3.24}$$

The prime candidate is a $25h_{50}^{+2}$eV tau neutrino. Note that this mass is well within the experimental bounds for the tau neutrino mass, and also that many particle

physics models – in particular those which lead to neutrino oscillations – predict masses of this order of magnitude.

3.4. Survey of Models

Any theory of structure formation must specify both the source of fluctuations and the composition of the dark matter. The reader is warned that the model called the "CDM Model" is a model with CDM **AND** perturbations generated by quantum fluctuations during a hypothetical period of inflation.

Inflation-based models were the first to be considered in quantitative detail, initially assuming a HDM-dominated Universe. Almost immediately, however, contradictions with basic observations appeared[54] (see, however, Ref. 214 for an opposing point of view).

The problem of HDM-based inflationary models is related to neutrino free streaming[55]. The primordial perturbations in this theory are dark matter fluctuations, but because of the large velocity of the dark matter particles, the inhomogeneities are washed out on all scales below the neutrino free streaming length $\lambda_j^c(t)$,

$$\lambda_j^c(t) \sim v(t) z(t) t \,, \tag{3.25}$$

which is the comoving distance the particles move in one Hubble expansion time. Since the neutrino velocity $v(t)$ and the redshift $z(t)$ both scale as $a(t)^{-1}$, the free streaming length decreases as

$$\lambda_j^c(t) \sim t^{-1/3} \tag{3.26}$$

after t_{eq} (before t_{eq} the radiation pressure dominates). Hence, in an inflationary HDM model all perturbations on scales λ smaller than the maximal value of $\lambda_j^c(t)$ are erased. The critical scale $\lambda_j^{\max}$ is given by the value of $\lambda_j^c(t)$ at the time when the neutrinos become non-relativistic which is in turn determined by the neutrino

mass m_ν. The result is

$$\lambda_j^{\max} \simeq 30\,\mathrm{Mpc}\,\left(\frac{m_\nu}{25\mathrm{eV}}\right)^{-2}\,, \tag{3.27}$$

a scale much larger than the mean separation of galaxies and clusters. Since we observe galaxies outside of large-scale structures, this model is in blatant disagreement with observations.

Inflation-based models are hence only viable if (at least a substantial fraction of) the dark matter is cold. Such models have become known as "CDM models", and are to a first approximation rather successful at predicting the clustering properties of galaxies and galaxy clusters[56]. There are many parameters in CDM models: the amplitude of the density perturbations, the power of the spectrum (see Section 4), the value of Ω, the fraction Ω_B of baryons, to mention some of the main ones. It is also possible to add a small fraction Ω_v of hot dark matter (yielding a class of so-called "Mixed Dark Matter" models).

Topological defect models were first developed in the context of CDM. Theories based on cosmic strings[57–59] or on global textures[45,60] have also been fairly successful in explaining observations (again to a first approximation).

It is important to note that if perturbations are seeded by long-lived topological defects (e.g., cosmic strings), then the above arguments against hot dark matter disappear[61,62]. The seed perturbations can survive neutrino free streaming as long as the seeds remain present for many Hubble expansion times. If we consider a comoving scale λ much smaller than $\lambda_j^{\max}$ of (3.27), then a dark matter perturbation will begin to grow about the seed fluctuations at a time $t(\lambda)$ when

$$\lambda_j^c(t(\lambda)) = \lambda\,. \tag{3.28}$$

Cosmic string based hot dark matter models have also been successful at explaining the qualitative features of observations[63].

4. Basics of Structure Formation

In the structure formation models mentioned in the previous section, small amplitude seed perturbations are predicted to arise due to particle physics effects in the very early Universe. They then grow by gravitational instability to produce the cosmological structures we observe today. In order to be able to make the connection between particle physics and observations, it is important to understand the gravitational evolution of fluctuations. This section will introduce the basic concepts of this topic. We begin, however, with an overview of some of the relevant data.

4.1. Survey of Data

It is length scales corresponding to galaxies and larger which are of greatest interest in cosmology when attempting to find an imprint of the primordial fluctuations produced by particle physics. On these scales, gravitational effects are assumed to be dominant, and the fluctuations are not too far from the linear regime. On smaller scales, nonlinear gravitational and hydrodynamical effects determine the final state and mask the initial perturbations.

To set the scales, consider the mean separation of galaxies, which is about $5h^{-1}$ Mpc[64], and that of Abell clusters which is around $25h^{-1}$ Mpc[65]. The largest coherent structures seen in current redshift surveys have a length of about $100h^{-1}$ Mpc, the recent detections of CMB anisotropies probe the density field on length scales of about 10^3h^{-1} Mpc, and the present horizon corresponds to a distance of about $3 \cdot 10^3h^{-1}$ Mpc.

Galaxies are gravitationally bound systems containing billions of stars. They are non-randomly distributed in space. A quantitative measure of this non-randomness is the "two-point correlation function" $\xi_2(r)$ which gives the excess probability of finding a galaxy at a distance r from a given galaxy:

$$\xi_2(r) =< \frac{n(r) - n_0}{n_0} > . \tag{4.1}$$

Here, n_0 is the average number density of galaxies, and $n(r)$ is the density of galaxies a distance r from a given one. The pointed braces stand for ensemble averaging.

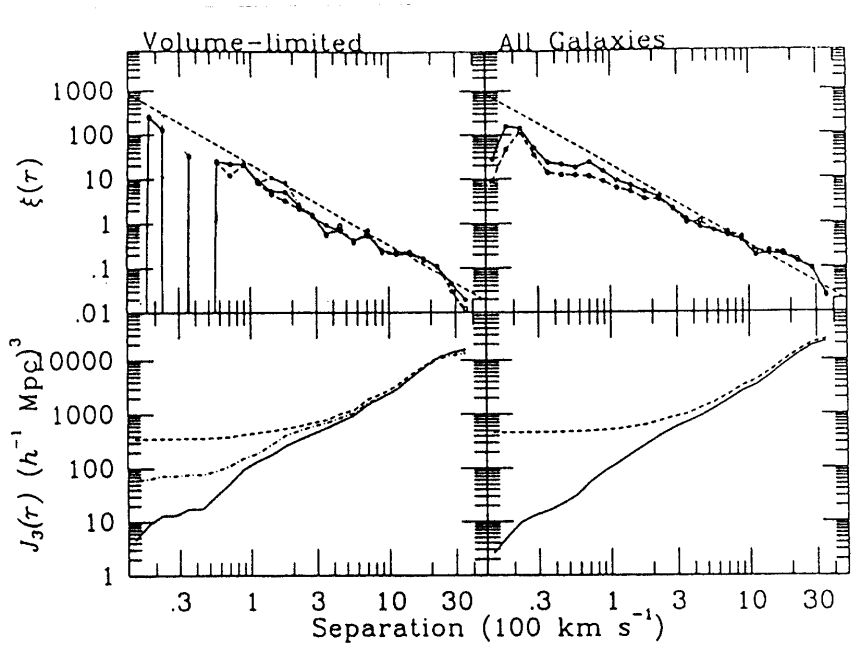

Figure 15: Recent observational results for the two point correlation function of IRAS galaxies, in both a volume limited subsample and a complete galaxy sample (see Ref. 66 which explains the meaning of the measure $J_3(r)$).

Recent observational results from a redshift survey of IRAS (infrared) galaxies yields reasonable agreement[66] with a form (see Fig. 15)

$$\xi_2(r) \simeq \left(\frac{r_0}{r}\right)^{\gamma} \tag{4.2}$$

with scaling length $r_0 \simeq 5h^{-1}$ Mpc and power $\gamma \simeq 1.8$. A theory of structure formation must explain both the amplitude and the slope of this correlation function.

Galaxies do not all have the same mass. There are more smaller galaxies than large ones (our galaxy is a large one). The distribution of galaxy masses is given by the "galaxy mass function" $n(M)$, where $n(M)dM$ is the number density of

galaxies in the mass range $[M,\, M + dM]$. Since we can only measure luminosity but not mass, the observable measure of the galaxy distribution is $\phi(L)$, the "galaxy luminosity function." Now, $\phi(L)dL$ is the number density of galaxies with luminosity L in the interval $[L,\, L + dL]$. Recent results for $\phi(L)$ from the IRAS galaxy survey are reproduced in Fig. 16[67].

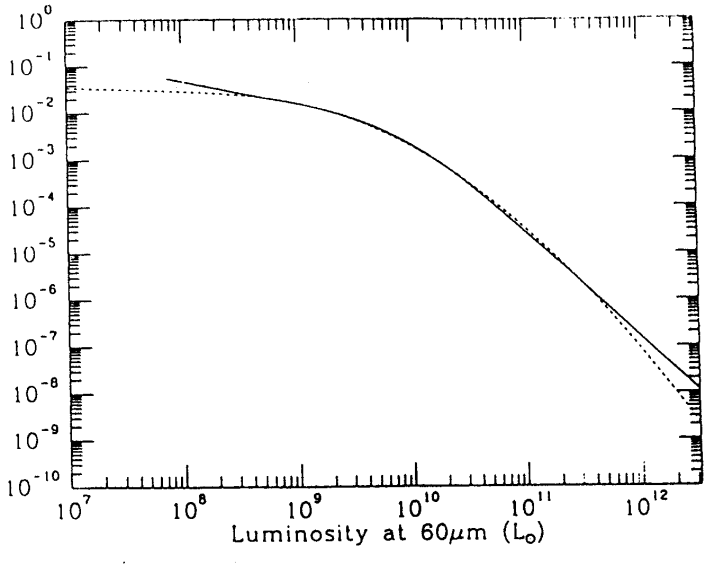

Figure 16: The luminosity function of IRAS galaxies. The vertical axis is $\phi(L)$, the horizontal axis is luminosity L in units of solar luminosity. The solid curve represents the data, the dashed curve is a fit to a theoretical model (see Ref. 67).

Theories must also be able to explain the internal mass distribution of galaxies which can be inferred from the galaxy rotation curves (see Fig. 14). According to the virial theorem, the velocity $v(r)$ at a distance r from the center of a galaxy is determined by the mass $M(r)$ inside of r:

$$\frac{mv^2}{r} = G\,\frac{mM(r)}{r^2} \tag{4.3}$$

where m is a test mass. Hence

$$v(r) = \left(G\,\frac{M(r)}{r}\right)^{1/2} . \tag{4.4}$$

A constant rotation velocity implies $M(r) \sim r$ and hence a density profile $\rho(r)$ of

$$\rho(r) \sim r^{-2}\,. \tag{4.5}$$

This condition puts constraints on the possible composition of the galactic dark matter.

An Abell cluster[68] is a region in space with greater than fifty bright galaxies in a sphere of radius $1.5h^{-1}$ Mpc, i.e., a region with a very high overdensity of galaxies. Observations indicate that Abell clusters are not distributed randomly in space. The cluster two point correlation function $\xi_c(r)$ has a form similar to (4.2)[65,69]:

$$\xi_c(r) \simeq \left(\frac{r_0}{r}\right)^{\gamma} \tag{4.6}$$

with $r_0 \simeq 15h^{-1}$ Mpc and $\gamma \simeq 2$ (see Fig. 17 which is taken from a recent analysis of rich clusters of galaxies selected from the APM Galaxy Survey[69]).

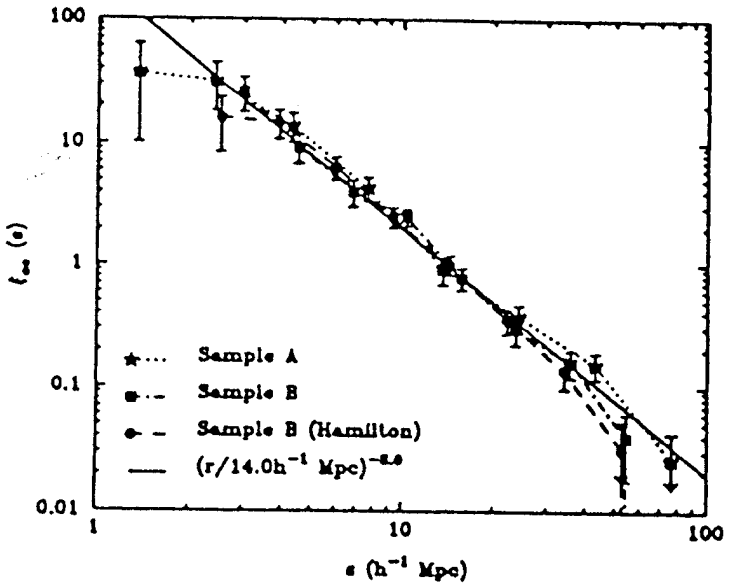

Figure 17: The two point correlation function ξ_c (denoted by ξ_{cc} in the figure) of clusters of galaxies

drawn from the APM galaxy survey as a function of their separation s. Results for two different samples of clusters are given.

There are several remarkable features about the clustering of galaxies and galaxy clusters. First, there is evidence for universality of the functional form of $\xi(r)$; its slope is about 2 for both populations. Secondly, relative to their respective mean separations, galaxies are more clustered than galaxy clusters. This can be explained by the action of gravity. Gravity has had longer to act on the scales of galaxies than on that of clusters, and has hence amplified the galaxy correlation function relative to that of clusters.

There is a wide spread of cluster masses which can be described by the cluster multiplicity function $\Phi_c(n)$, where $\Phi_c(n)dn$ is the number density of clusters containing between n and $n + dn$ galaxies. Fig. 18 is a sketch of the observed cluster multiplicity function taken from Ref. 70. Note that the cluster mass function inferred from Fig. 18 and the galaxy mass function deduced from the galaxy luminosity function of Fig. 16 match up quite well at a mass of about $10^{12}M_{\odot}$. Below this mass, the objects are well defined dynamical entities whereas for larger masses they are composed of fragments. A reason for the difference may be due to the fact that clouds of more than $10^{12}M_{\odot}$ cannot cool without fragmenting[71].

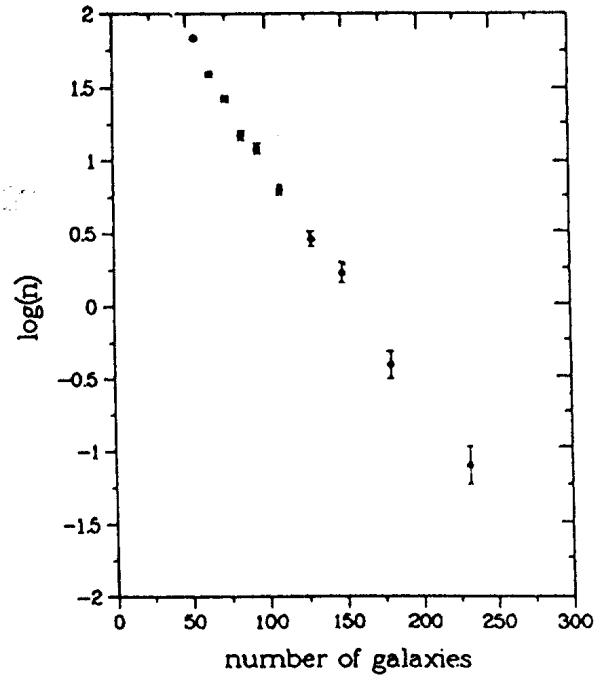

Figure 18: The cluster multiplicity function $\Phi_c(n)$ (denoted as n on the figure) as a function of the number n (horizontal axis).

On scales larger than galaxy clusters there is not at present a clear mathematical description of structure. Many galaxy redshift surveys have discovered coherent filamentary and planar structures and voids on scales of up to $100h^{-1}$ Mpc[14,72–75]. For example, the astronomers working on the "Center for Astrophysics" redshift survey[14] have analyzed many adjacent slices of the northern celestial sphere. For all galaxies above a limiting magnitude of 15.5 they measured the redshifts z. Fig. 19 is a sketch of redshift versus angle α in the sky for one slice. The second direction in the sky has been projected onto the $\alpha - z$ plane. The most prominent feature is the band of galaxies at a distance of about $100h^{-1}$ Mpc. This band also appears in neighboring slices and is therefore presumably part of a planar density enhancement of comoving planar size of at least $(50 \times 100) \times h^{-2}$ Mpc^2. This structure is often called the "great wall." It is a challenge for theories of structure formation to explain both the observed scale and topology of the galaxy distribution.

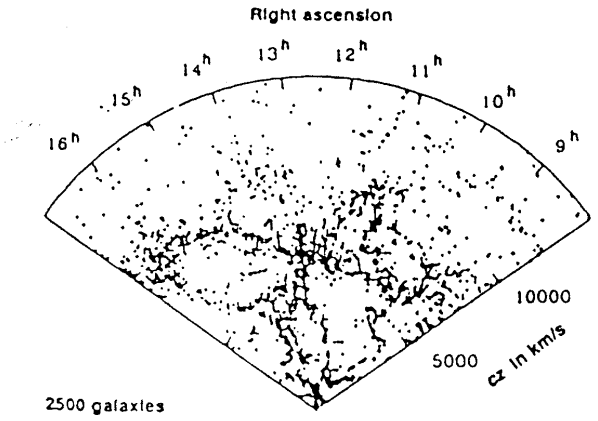

Figure 19: Results from the CFA redshift survey. Radial distance gives the redshift of galaxies, the angular distance corresponds to right ascension. The results from several slices of the sky (at different declinations) have been projected into the same cone.

Until 1992 there was little evidence for any convergence of the galaxy distribution towards homogeneity. Each new survey led to the discovery of new coherent structures in the Universe on a scale comparable to that of the survey. In 1992, preliminary results of a much deeper redshift survey were announced[15)] which for the first time found no new coherent structures on scales larger than $100h^{-1}$ Mpc. This is the first direct evidence for the cosmological principle from optical surveys (the isotropy of the CMB has for a long time been a strong point in its support).

In summary, a lot of data from optical and infrared galaxies alone are currently available, and new data are being collected at a rapid rate. The observational constraints on theories of structure formation are becoming tighter. A lot of theoretical work is needed in order to allow for detailed comparisons between theory and observations.

4.2. Gravitational Instability

In this article we only discuss theories in which structures grow by gravitational accretion. The basic mechanism is illustrated in Fig. 20. Consider first a flat space-time background. A density perturbation with $\delta\rho > 0$ will then give rise to an excess gravitational attractive force F acting on the surrounding matter. This force is proportional to $\delta\rho$, and will hence lead to exponential growth of the perturbation since

$$\delta\ddot{\rho} \sim F \sim \delta\rho \Rightarrow \delta\rho \sim \exp(\alpha t) \tag{4.7}$$

with some constant α.

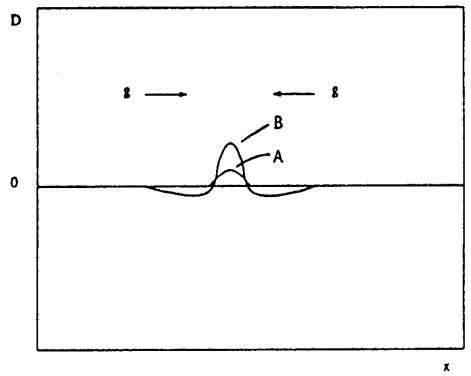

Figure 20: Sketch of the gravitational instability mechanism. The vertical axis is the density perturbation D as a function of a line in space (x). A small initial overdensity (A) will cause a gravitational acceleration g towards it, which will lead to an increase in the perturbation (B). Note that in general underdense regions develop in addition to the growing overdense areas.

In an expanding background space-time, the acceleration is damped by the expansion. If $r(t)$ is the physcial distance of a test particle from the perturbation,

then on a scale r

$$\delta\ddot{\rho} \sim F \sim \frac{\delta\rho}{r^2(t)}\,, \tag{4.8}$$

which results in power-law increase of $\delta\rho$. The goal of this subsection is to discuss the growth rates of inhomogeneities in more detail (see e.g. Refs. 76 and 77 for modern reviews).

Because of our assumption that all perturbations start out with a small amplitude, we can linearize the equations for gravitational fluctuations. The analysis is then greatly simplified by going to Fourier space in which all modes $\delta(\underline{k})$ decouple. We expand the fractional density contrast $\delta(\underline{x})$ as follows:

$$\delta(\underline{x}) = \frac{\delta\rho(x)}{\rho} = (2\pi)^{-3/2} V^{1/2} \int d^3k\, e^{i\underline{k}\cdot\underline{x}} \delta(\underline{k})\,, \tag{4.9}$$

where V is a cutoff volume which disappears from all physical observables.

The "power spectrum" $P(k)$ is defined by

$$P(k) = < |\delta(k)|^2 >\,, \tag{4.10}$$

where the braces denote an ensemble average (in most structure formation models, the generation of perturbations is a stochastic process, and hence observables can only be calculated by averaging over the ensemble. For observations, the braces can be viewed as an angular average).

The physical measure of mass fluctuations on a length scale λ is the r.m.s. mass fluctuation $\delta M/M(\lambda)$ on this scale. It is determined by the power spectrum in the following way. We pick a center $\underline{x}_0$ of a sphere $B_\lambda(x_0)$ of radius λ and calculate

$$\left|\frac{\delta M}{M}\right|^2 (\underline{x}_0,\, \lambda) = \left| \int\limits_{B_\lambda(\underline{x}_0)} d^3x \delta(\underline{x})\, \frac{1}{V(B_\lambda)} \right|^2, \tag{4.11}$$

where $V(B_\lambda)$ is the volume of the sphere. Inserting the Fourier decomposition

(4.3) we obtain

$$|\frac{\delta M}{M}|^2(\underline{x}_0,\,\lambda) = \frac{V}{(2\pi)^3}\,\frac{1}{V(B_\lambda)^2}\int\limits_{B_\lambda(0)} d^3x_1 \int\limits_{B_\lambda(0)} d^3x_2 \int d^4k_1 d^4k_2 e^{i(\underline{k}_1-\underline{k}_2)\cdot\underline{x}_0} \\ e^{i\underline{k}_1\cdot\underline{x}_1} e^{-i\underline{k}_2\cdot\underline{x}_2}\delta(\underline{k}_1)^*\,\delta(\underline{k}_2)\,. \tag{4.12}$$

Taking the average value of this quantity over all $\underline{x}_0$ yields

$$< \left(\frac{\delta M}{M}\right)^2(\lambda) >= \int d^3k W_k(\lambda)|\delta(\underline{k}|^2 \tag{4.13}$$

with a window function $W_k(\lambda)$ with the following properties

$$W_k(\lambda)\begin{cases} \simeq 1 & k < k_\lambda = 2\pi/\lambda \\ \simeq 0 & k > k_\lambda\ . \end{cases} \tag{4.14}$$

Therefore the r.m.s. mass perturbation on a scale λ becomes

$$< \left|\frac{\delta M}{M}(\lambda)\right|^2 > \sim k_\lambda^3 P(k_\lambda)\,. \tag{4.15}$$

Astronomers usually assume that $P(k)$ grows as a power of k:

$$P(k) \sim k^n\,, \tag{4.16}$$

n being called the index of the power spectrum. For $n = 1$ we get the so-called Harrison-Zel'dovich scale invariant spectrum[78].

Both inflationary Universe and topological defect models of structure formation predict a roughly scale invariant spectrum. The distinguishing feature of this spectrum is that the r.m.s. mass perturbations are independent of the scale k when measured at the time $t_H(k)$ when the associated wavelength is equal to the Hubble radius, i.e., when the scale "enters" the Hubble radius. Let us derive this

fact for the scales entering during the matter dominated epoch. The time $t_H(k)$ is determined by

$$k^{-1}a(t_H(k)) = t_H(k) \tag{4.17}$$

which leads to

$$t_H(k) \sim k^{-3}\,. \tag{4.18}$$

According to the linear theory of cosmological perturbations discussed in the following subsection, the mass fluctuations increase as $a(t)$ for $t > t_{eq}$. Hence

$$\frac{\delta M}{M}(k, t_H(k)) = \left(\frac{t_H(k)}{t}\right)^{2/3} \frac{\delta M}{M}\,(k,t) \sim \text{const}\,, \tag{4.19}$$

since the first factor scales as k^{-2} and – using (4.15) and inserting $n = 1$ – the second as k^2.

4.3. Newtonian Theory of Cosmological Perturbations

The Newtonian theory of cosmological perturbations is an approximate analysis which is valid on wavelengths λ much smaller than the Hubble radius t and for negligible pressure p, i.e., $p \ll \rho$. It is based on expanding the hydrodynamical equations about a homogeneous background solution.

The starting points are the continuity, Euler and Poisson equations

$$\dot{\rho} + \underline{\nabla}(\rho \underline{v}) = 0 \tag{4.20}$$

$$\dot{\underline{v}} + (\underline{v} \cdot \underline{\nabla})\underline{v} = -\underline{\nabla}\phi - \frac{1}{\rho}\,\underline{\nabla} p \tag{4.21}$$

$$\nabla^2 \phi = 4\pi G \rho \tag{4.22}$$

for a fluid with energy density ρ, pressure p, velocity $\underline{v}$ and Newtonian gravitational potential ϕ, written in terms of physical coordinates $(t, \underline{r})$.

The transition to an expanding space is made by introducing comoving coordinates $\underline{x}$ and peculiar velocity $\underline{u} = \dot{\underline{x}}$:

$$\underline{r} = a(t)\underline{x} \tag{4.23}$$

$$\underline{v} = \dot{a}(t)\underline{x} + a(t)\underline{u}\,. \tag{4.24}$$

The first term on the right hand side of (4.24) is the expansion velocity.

The perturbation equations are obtained by linearizing Equations (4.20-22) about a homogeneous background solution $\rho = \bar{\rho}(t)$, $p = 0$ and $\underline{u} = 0$. Using the definition

$$\delta \equiv \frac{\delta\rho}{\rho}\,, \tag{4.25}$$

the linearization ansatz can be written

$$\rho(\underline{x}, t) = \bar{\rho}(t)(1 + \delta(\underline{x}, t))\,. \tag{4.26}$$

If we consider adiabatic perturbations (no entropy density variations), then after some algebra the linearized equations become

$$\dot{\delta} + \nabla \cdot \underline{u} = 0\,, \tag{4.27}$$

$$\dot{\underline{u}} + 2\frac{\dot{a}}{a}\underline{u} = -a^2(\nabla\delta\phi + c_s^2\nabla\delta) \tag{4.28}$$

and

$$\nabla^2\delta\phi = 4\pi G\bar{\rho}a^2\delta\,, \tag{4.29}$$

with the speed of sound c_s given by

$$c_s^2 = \frac{\partial p}{\partial\rho}\,. \tag{4.30}$$

The two first order equations (4.27) and (4.28) can be combined to yield a single

second order differential equation for δ. With the help of (4.29) this equation reads

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}\delta - \frac{c_s^2}{a^2}\nabla^2\delta = 0 \tag{4.31}$$

which in Fourier space becomes

$$\ddot{\delta}_{\underline{k}} + 2H\dot{\delta}_{\underline{k}} + \left(\frac{c_s^2 k^2}{a^2} - 4\pi G\bar{\rho}\right)\delta_{\underline{k}} = 0\,. \tag{4.32}$$

Here, $H(t)$ as usual denotes the expansion rate, and $\delta_{\underline{k}}$ stands for $\delta(\underline{k})$.

Already a quick look at Equation (4.32) reveals the presence of a distinguished scale for cosmological perturbations, the Jeans length

$$\lambda_J = \frac{2\pi}{k_J} \tag{4.33}$$

with

$$k_J^2 = \left(\frac{k}{a}\right)^2 = \frac{4\pi G\bar{\rho}}{c_s^2}\,. \tag{4.34}$$

On length scales larger than λ_J, the spatial gradient term is negligible, and the term linear in δ in (4.32) acts like a negative mass square quadratic potential with damping due to the expansion of the Universe, in agreement with the intuitive analysis leading to (4.7) and (4.8). On length scales smaller than λ_J, however, (4.32) becomes a damped harmonic oscillator equation and perturbations on these scales decay.

For $t > t_{eq}$ and for $\lambda \gg \lambda_J$, Equation (4.32) becomes

$$\ddot{\delta}_k + \frac{4}{3t}\,\dot{\delta}_k - \frac{2}{3t^2}\,\delta_k = 0 \tag{4.35}$$

and has the general solution

$$\delta_k(t) = c_1 t^{2/3} + c_2 t^{-1}\,. \tag{4.36}$$

This demonstrates that for $t > t_{eq}$ and $\lambda \gg \lambda_J$, the dominant mode of perturbations increases as $a(t)$, a result we already used in the previous subsection (see (4.19)).

For $\lambda \ll \lambda_J$ and $t > t_{eq}$, Equation (4.32) becomes

$$\ddot{\delta}_k + 2H\dot{\delta}_k + c_s^2 \left(\frac{k}{a}\right)^2 \delta_k = 0\,, \tag{4.37}$$

and has solutions corresponding to damped oscillations:

$$\delta_k(t) \sim a^{-1/2}(t) \exp\{\pm i c_s k \int dt' a(t')^{-1}\}\,. \tag{4.38}$$

As an important application of the Newtonian theory of cosmological perturbations, let us compare sub-horizon scale fluctuations in a baryon-dominated Universe ($\Omega = \Omega_B = 1$) and in a CDM-dominated Universe with $\Omega_{CDM} = 0.9$ and $\Omega = 1$. We consider scales which enter the Hubble radius at about t_{eq}.

In the initial time interval $t_{eq} < t < t_{rec}$, the baryons are coupled to the photons. Hence, the baryonic fluid has a large pressure p_B

$$p_B \simeq p_r = \frac{1}{3}\,\rho_r\,. \tag{4.39}$$

Hence, the speed of sound is relativistic

$$c_s \simeq \left(\frac{p_r}{\rho_m}\right)^{1/2} = \frac{1}{\sqrt{3}}\,\left(\frac{\rho_r}{\rho_m}\right)^{1/2}\,. \tag{4.40}$$

The value of c_s slowly decreases in this time interval, attaining a value of about $1/10$ at t_{rec}. From (4.34) it follows that the Jeans mass M_J, the mass inside a sphere of radius λ_J, increases until t_{rec} when it reaches its maximal value M_J^{max}

$$M_J^{max} = M_J(t_{rec}) = \frac{4\pi}{3}\,\lambda_J(t_{rec})^3 \bar{\rho}(t_{rec}) \sim 10^{17}(\Omega h^2)^{-1/2} M_\odot\,. \tag{4.41}$$

At the time of recombination, the baryons decouple from the radiation fluid. Hence, the baryon pressure p_B drops abruptly, as does the Jeans length (see (4.34)).

The remaining pressure p_B is determined by the temperature and thus continues to decrease as t increases. It can be shown that the Jeans mass continues to decrease after t_{rec}, starting from a value

$$M_J^-(t_{rec}) \sim 10^6 (\Omega h^2)^{-1/2} \, M_\odot \tag{4.42}$$

(where the superscript "$-$" indicates the mass immediately after t_{eq}.

In contrast, CDM has negligible pressure throughout the period $t > t_{eq}$ and hence experiences no Jeans damping. A CDM perturbation which enters the Hubble radius at t_{eq} with amplitude δ_i has an amplitude at t_{rec} given by

$$\delta_k^{CDM}(t_{rec}) \simeq \frac{a(t_{rec})}{a(t_{eq})} \, \delta_i \, , \tag{4.43}$$

whereas a perturbation with the same scale and initial amplitude in a baryon-dominated Universe is damped

$$\delta_k^{BDM}(t_{rec}) \simeq \left(\frac{a(t_{eq})}{a(t_{eq})} \right)^{-1/2} \delta_i \, . \tag{4.44}$$

In order for the perturbations to have the same amplitude today, the initial size of the inhomogeneity must be much larger in a BDM-dominated Universe than in a CDM-dominated one:

$$\delta_k^{BDM}(t_{eq}) \simeq \left(\frac{z(t_{eq})}{z(t_{eq})} \right)^{3/2} \delta_k^{CDM}(t_{eq}) \, . \tag{4.45}$$

For $\Omega = 1$ and $h = 1/2$ the enhancement factor is about 30.

In a CDM-dominated Universe the baryons experience Jeans damping, but after t_{rec} the baryons quickly fall into the potential wells created by the CDM perturbations, and hence the baryon perturbations are proportional to the CDM inhomogeneities (see Fig. 21).

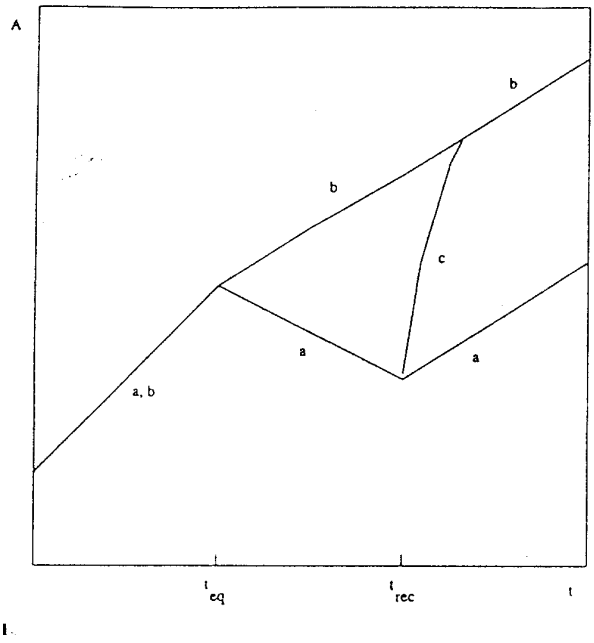

Figure 21: Comparison of the growth of fluctuations in baryon and CDM dominated Universes. The horizontal axis is time, the vertical axis is the density perturbation. The curve labelled by a describes the evolution of fluctuations in a BDM Universe. The growth of CDM perturbations follows the curve b, and curve c is a sketch of the time development of baryon perturbations in a CDM dominated Universe. To be specific, perturbations on a scale which enters the Hubble radius at t_{eq} are considered.

The above considerations, coupled with information about CMB anisotropies, can be used to rule out a model with $\Omega = \Omega_B = 1$. The argument goes as follows (see Section 7). For adiabatic fluctuations, the amplitude of CMB anisotropies on an angular scale ϑ is determined by the value of $\delta\rho/\rho$ on the corresponding length scale $\lambda(\vartheta)$ at t_{eq}:

$$\frac{\delta T}{T}(\vartheta) = \frac{1}{3}\,\frac{\delta\rho}{\rho}(\lambda(\vartheta),\, t_{eq})\,. \tag{4.46}$$

On scales of clusters we know that (for $\Omega = 1$ and $h = 1/2$)

$$\left(\frac{\delta\rho}{\rho}\right)_{CDM}(\lambda(\vartheta),\, t_{eq}) \simeq z(t_{eq})^{-1} \simeq 10^{-4}\,, \tag{4.47}$$

using the fact that today on cluster scales $\delta\rho/\rho \simeq 1$. The bounds on $\delta T/T$ on

small angular scales are

$$\frac{\delta T}{T} << (\vartheta) 10^{-4} \,, \tag{4.48}$$

consistent with the predictions for a CDM model, but inconsistent with those of a $\Omega = \Omega_B = 1$ model, according to which we would expect anisotropies of the order of 10^{-3}. This is yet another argument in support of the existence of nonbaryonic dark matter.

To conclude this subsection, let us briefly discuss Newtonian perturbations during the radiation-dominated epoch. We consider matter fluctuations with $c_s = 0$ in a smooth relativistic background. In this case, Equation (4.32) becomes

$$\ddot{\delta}_k + 2H\dot{\delta}_k - 4\pi G \bar{\rho}_m \delta_k = 0 \,, \tag{4.49}$$

where $\bar{\rho}_m$ denotes the average matter energy density. The Hubble expansion parameter obeys

$$H^2 = \frac{8\pi G}{3}(\bar{\rho}_m + \bar{\rho}_r) \,, \tag{4.50}$$

with $\bar{\rho}_r$ the background radiation energy density. For $t < t_{eq}$, $\bar{\rho}_m$ is negligible in both (4.49) and (4.50), and (4.49) has the general solution

$$\delta_k(t) = c_1 \log t + c_2 \,. \tag{4.51}$$

In particular, this result implies that CDM perturbations which enter the Hubble radius before t_{eq} have an amplitude which grows only logarithmically in time until t_{eq}. This is sometimes called the Meszaros effect.

4.4. Relativistic Theory of Cosmological Perturbations

On scales larger than the Hubble radius ($\lambda > t$) the Newtonian theory of cosmological perturbations obviously is inapplicable, and a general relativistic analysis is needed. On these scales, matter is essentially frozen in comoving coordinates. However, space-time fluctuations can still increase in amplitude.

In principle, it is straightforward to work out the general relativistic theory of linear fluctuations[79]. We linearize the Einstein equations

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{4.52}$$

(where $G_{\mu\nu}$ is the Einstein tensor associated with the space-time metric $g_{\mu\nu}$, and $T_{\mu\nu}$ is the energy-momentum tensor of matter) about an expanding FRW background ($g^{(0)}_{\mu\nu}$, $\varphi^{(0)}$):

$$\begin{aligned} g_{\mu\nu}(\underline{x}, t) &= g^{(0)}_{\mu\nu}(t) + h_{\mu\nu}(\underline{x}, t) \\ \varphi(\underline{x}, t) &= \varphi^{(0)}(t) + \delta\varphi(\underline{x}, t) \end{aligned} \tag{4.53}$$

and pick out the terms linear in $h_{\mu\nu}$ and $\delta\varphi$ to obtain

$$\delta G_{\mu\nu} \;=\; 8\pi G \delta T_{\mu\nu}\,. \tag{4.54}$$

In the above, $h_{\mu\nu}$ is the perturbation in the metric and $\delta\varphi$ is the fluctuation of the matter field φ. We have denoted all matter fields collectively by φ.

In practice, there are many complications which make this analysis highly nontrivial. The first problem is "gauge invariance"[80] Imagine starting with a homogeneous FRW cosmology and introducing new coordinates which mix $\underline{x}$ and t. In terms of the new coordinates, the metric now looks inhomogeneous. The inhomogeneous piece of the metric, however, must be a pure coordinate (or "gauge") artefact. Thus, when analyzing relativistic perturbations, care must be taken to factor out effects due to coordinate transformations.

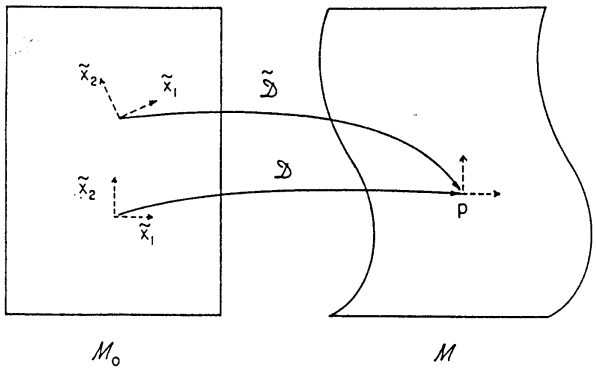

Figure 22: Sketch of how two choices of the mapping from the background space-time manifold $\mathcal{M}_0$ to the physical manifold $\mathcal{M}$ induce two different coordinate systems on $\mathcal{M}$.

The issue of gauge dependence is illustrated in Fig. 22. A coordinate system on the physical inhomogeneous space-time manifold $\mathcal{M}$ can be viewed as a mapping $\mathcal{D}$ of an unperturbed space-time $\mathcal{M}_0$ into $\mathcal{M}$. A physical quantity Q is a geometrical function defined on $\mathcal{M}$. There is a corresponding physical quantity ${}^{(0)}Q$ defined on $\mathcal{M}_0$. In the coordinate system given by $\mathcal{D}$, the perturbation δQ of Q at the space-time point $p \,\epsilon\, \mathcal{M}$ is

$$\delta Q(p) = Q(p) - {}^{(0)}Q\left(D^{-1}(p)\right). \tag{4.55}$$

However, in a second coordinate system $\tilde{\mathcal{D}}$ the perturbation is given by

$$\delta\tilde{Q}(p) = Q(p) - {}^{(0)}Q(\tilde{\mathcal{D}}^{-1}(p)). \tag{4.56}$$

The difference

$$\Delta Q(p) = \delta Q(p) - \delta\tilde{Q}(p) \tag{4.57}$$

is obviously a gauge artefact and carries no physical meaning.

There are various methods of dealing with gauge artefacts. The simplest and most physical approach is to focus on gauge invariant variables, i.e., combinations

of the metric and matter perturbations which are invariant under linear coordinate transformations.

The gauge invariant theory of cosmological perturbations is in principle straightforward, although technically rather tedious. In the following I will summarize the main steps and refer the reader to Ref. 11 for the details and further references (see also Ref. 81 for a pedagogical introduction and Refs. 82-87 for other approaches).

We consider perturbations about a spatially flat Friedmann-Robertson-Walker metric

$$ds^2 = a^2(\eta)(d\eta^2 - d\underline{x}^2) \tag{4.58}$$

where η is conformal time (related to cosmic time t by $a(\eta)d\eta = dt$). A scalar metric perturbation (see Ref. 88 for a precise definition) can be written in terms of four free functions of space and time:

$$\delta g_{\mu\nu} = a^2(\eta) \begin{pmatrix} 2\phi & -B_{,i} \\ -B_{,i} & 2(\psi\delta_{ij} + E_{,ij} \end{pmatrix} . \tag{4.59}$$

Scalar metric perturbations are the only perturbations which couple to energy density and pressure.

The next step is to consider infinitesimal coordinate transformations

$$x^{\mu'} = x^\mu + \xi^\mu \tag{4.60}$$

which preserve the scalar nature of $\delta g_{\mu\nu}$ and to calculate the induced transformations of ϕ, ψ, B and E. Then we find invariant combinations to linear order. (Note that there are in general no combinations which are invariant to all orders[89].) After some algebra, it follows that

$$\begin{aligned} \Phi &= \phi + a^{-1}[(B - E')a]' \\ \Psi &= \psi - \frac{a'}{a}(B - E') \end{aligned} \tag{4.61}$$

are two invariant combinations. In the above, a prime denotes differentiation with respect to η.

There are various methods to derive the equations of motion for gauge invariant variables. Perhaps the simplest way[11] is to consider the linearized Einstein equations (4.54) and to write them out in the longitudinal gauge defined by

$$B = E = 0 \tag{4.62}$$

and in which $\Phi = \phi$ and $\Psi = \psi$, to directly obtain gauge invariant equations.

For several types of matter, in particular for scalar field matter, the perturbation of $T_{\mu\nu}$ has the special property

$$\delta T^i_j \sim \delta^i_j \tag{4.63}$$

which imples $\Phi = \Psi$. Hence, the scalar-type cosmological perturbations can in this case be described by a single gauge invariant variable. The equation of motion takes the form[90,9,10]

$$\dot{\xi} = O\left(\frac{k}{aH}\right)^2 H\xi \tag{4.64}$$

where

$$\xi = \frac{2}{3}\,\frac{H^{-1}\dot{\Phi} + \Phi}{1+w} + \Phi\,. \tag{4.65}$$

The variable $w = p/\rho$ (with p and ρ background pressure and energy density respectively) is a measure of the background equation of state. In particular, on scales larger than the Hubble radius, the right hand side of (4.64) is negligible, and hence ξ is constant.

The result that $\dot{\xi} = 0$ is a very powerful one. Let us first imagine that the equation of state of matter is constant, *i.e.*, $w = \text{const}$. In this case, $\dot{\xi} = 0$ implies

$$\Phi(t) = \text{const}\,, \tag{4.66}$$

i.e., this gauge invariant measure of perturbations remains constant outside the Hubble radius.

Next, consider the evolution of Φ during a phase transition from an initial phase with $w = w_i$ to a phase with $w = w_f$. Long before and after the transition, Φ is constant because of (4.66), and hence $\dot{\xi} = 0$ becomes

$$\frac{\Phi}{1+w} + \Phi = \text{const}\,, \tag{4.67}$$

In order to make contact with matter perturbations and Newtonian intuition, it is important to remark that, as a consequence of the Einstein constraint equations, at Hubble radius crossing Φ is a measure of the fractional density fluctuations:

$$\Phi(k, t_H(k)) \sim \frac{\delta\rho}{\rho}\,(k,\, t_H(k))\,. \tag{4.68}$$

(Note that the latter quantity is approximately gauge invariant on scales smaller than the Hubble radius).

5. Inflationary Universe Scenarios

5.1. Preliminaries

Cosmological inflation[30] is a period in time during which the Universe is expanding exponentially, *i.e.*,

$$a(t) = e^{tH} \tag{5.1}$$

with constant Hubble expansion rate H. From the FRW equations (2.7) and (2.9) it follows that the condition for inflation (in the context of Einstein gravity in a spatially flat Universe) is an equation of state for matter with

$$p = -\rho\,, \tag{5.2}$$

which neccessitates abandoning a description of matter in terms of an ideal gas.

As was indicated in Section 3.1, it is possible to achieve inflation if matter is described in terms of scalar fields, provided that at some period

$$\begin{aligned} \dot{\varphi}^2 &\ll V(\varphi) \\ (\nabla\varphi)^2 &\ll V(\varphi) \end{aligned} \tag{5.3}$$

(see Eqs. (3.9) which give the equation of state for scalar field matter).

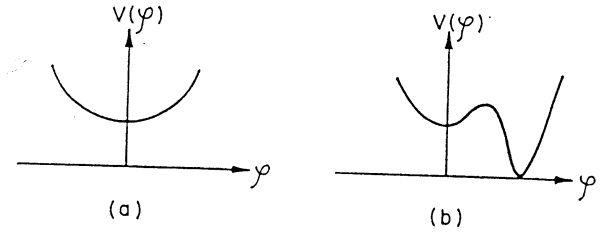

Figure 23: A sketch of two potentials which can give rise to inflation.

Two examples which can give inflation are shown in Fig. 23. In (a), inflation occurs at the stable fixed point $\varphi(\underline{x}, t_i) = 0 = \dot{\varphi}(\underline{x}, t_i)$. However, this model is ruled out by observation: the inflationary phase has no ending. $V(0)$ acts as a permanent nonvanishing cosmological constant. In (b), a finite period of inflation can arise if $\varphi(\underline{x})$ is trapped at the local minimum $\varphi = 0$ with $\dot{\varphi}(\underline{x}) = 0$. However, in this case $\varphi(\underline{x})$ can make a sudden transition at some time $t_R > t_i$ through the potential barrier and move to $\varphi(\underline{x}) = a$. Thus, for $t_i < t < t_R$ the Universe expands exponentially, whereas for $t > t_R$ the contribution of φ to the expansion of the Universe vanishes and we get the usual FRW cosmology. There are three obvious questions: why does the field start out at $\varphi = 0$, how does the transition occur and why should the scalar field have $V(\varphi) = 0$ at the global minimum? In the following section the first two questions will be addressed. The third question is part of the cosmological constant problem for which there is as yet no convincing explanation. Before studying the dynamics of the phase transition, we need to digress and discuss finite temperature effects.

5.2. Finite Temperature Field Theory

The evolution of particles in vacuum and in a thermal bath are very different. Similarly, the evolution of fields changes when coupled to a thermal bath. Under certain conditions, the changes may be absorbed in a temperature dependent potential, the finite temperature effective potential. Here, a heuristic derivation of this potential will be given. The reader is referred to Ref. 8 or to the original articles[91] for the actual derivation.

We assume that the scalar field $\varphi(\underline{x}, t)$ is coupled to a thermal bath which is represented by a second scalar field $\psi(\underline{x}, t)$ in thermal equilibrium. The Lagrangian for φ is

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi \, - \, V(\varphi) \, - \, \frac{1}{2}\hat{\lambda}\varphi^2\psi^2 \, , \tag{5.4}$$

where $\hat{\lambda}$ is a coupling constant. The action from which the equations of motion are derived is

$$S \, = \, \int d^4x \, \sqrt{-g}\mathcal{L} \tag{5.5}$$

where g is the determinant of the metric (2.2). The resulting equation of motion for $\varphi(\underline{x}, t)$ is

$$\ddot{\varphi} + 3H\dot{\varphi} \, - a^{-2} \bigtriangledown^2 \varphi \, = \, -V'(\varphi) - \hat{\lambda}\psi^2\varphi \, . \tag{5.6}$$

If ψ is in thermal equilibrium, we may replace ψ^2 by its thermal expectation value $< \psi^2 >_T$. Now,

$$< \psi^2 >_T \sim \, T^2 \tag{5.7}$$

which can be seen as follows: in thermal equilibrium, the energy density of ψ equals that of one degree of freedom in the thermal bath. In particular, the potential energy density $V(\psi)$ of ψ is of that order of magnitude. Let

$$V(\psi) \, = \, \lambda_\psi\psi^4 \tag{5.8}$$

with a coupling constant λ_ψ which we take to be of the order 1 (if λ_ψ is too

small, ψ will not be in thermal equilibrium). Since the thermal energy density is proportional to T^4, (5.7) follows. (5.6) can be rewritten as

$$\ddot{\psi} + 3H\dot{\varphi} - a^{-2} \bigtriangledown^2 \varphi = -V_T'(\varphi), \tag{5.9}$$

where

$$V_T(\varphi) = V(\varphi) + \frac{1}{2}\hat{\lambda}T^2\varphi^2 \tag{5.10}$$

is called the finite temperature effective potential. Note that in (5.10), $\hat{\lambda}$ has been rescaled to absorb the constant of proportionality in (5.7).

These considerations will now be applied to Example A, a scalar field model with potential

$$V(\varphi) = \frac{1}{4}\lambda(\varphi^2 - \sigma^2)^2 \tag{5.11}$$

(σ is called the scale of symmetry breaking). The finite temperature effective potential becomes (see Fig. 24)

$$V_T(\varphi) = \frac{1}{4}\lambda\varphi^4 - \frac{1}{2}\left(\lambda\sigma^2 - \hat{\lambda}T^2\right)\varphi^2 + \frac{1}{4}\lambda\sigma^4 . \tag{5.12}$$

For very high temperatures, the effective mass term is positive and hence the energetically favorable state is $<\varphi>= 0$. For very low temperatures, on the other hand, the mass term has a negative sign which leads to spontaneous symmetry breaking. The temperature at which the mass term vanishes defines the critical temperature T_c

$$T_c = \hat{\lambda}^{-1/2}\lambda^{1/2}\sigma . \tag{5.13}$$

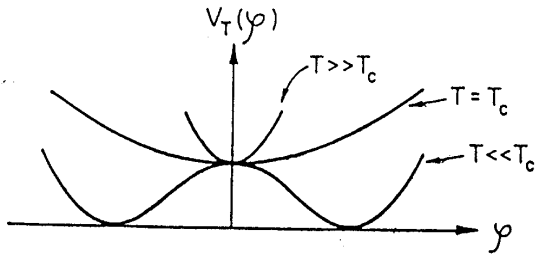

Figure 24: The finite temperature effective potential for Example A.

As Example B, consider a theory with potential

$$V(\varphi) = \frac{1}{4}\varphi^4 - \frac{1}{3}(a+b)\varphi^3 + \frac{1}{2}ab\varphi^2 \tag{5.14}$$

with $\frac{1}{2}a > b > 0$. The finite temperature effective potential is obtained by adding $\frac{1}{2}\hat{\lambda}T^2\varphi^2$ to the right hand side of (5.14). $V_T(\varphi)$ is sketched in Fig. 25 for various values of T. The critical temperature T_c is defined as the temperature when the two minima of $V_T(\varphi)$ become degenerate.

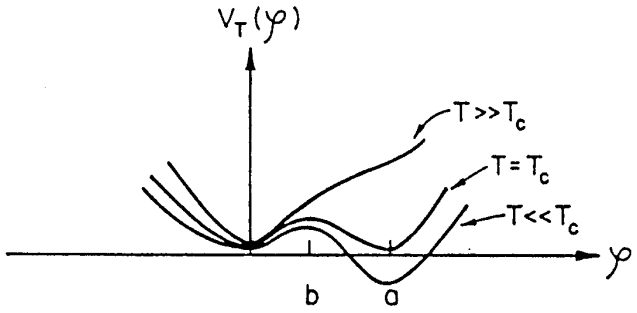

Figure 25: The finite temperature effective potential for Example B.

It is important to note that the use of finite temperature effective potential methods is only legitimate if the system is in thermal equilibrium. This point was stressed in Refs. 92 and 93, although the conclusion should be obvious from

the derivation given above. To be more precise, we require the ψ field to be in thermal equilibrium and the coupling constant $\hat{\lambda}$ of (5.4) which mediates the energy exchange between the φ and ψ fields to be large. However, as shown in Chapter 4, in inflationary Universe models, the observational constraints stemming from the amplitude of the primordial energy density fluctuation spectrum force the self coupling constant λ of φ to be extremely small. Since at one loop order, the interaction term $\frac{1}{2}\hat{\lambda}\varphi^2\psi^2$ induces contributions to λ, it is unnatural to have λ very small and $\hat{\lambda}$ unsuppressed. Hence, in many inflationary Universe models - in particular in new inflation[94] and in chaotic inflation[92] - finite temperature effective potential methods are inapplicable.

5.3. Phase Transitions

The temperature dependence of the finite temperature effective potential in quantum field theory leads to phase transitions in the very early Universe. These transitions are either first or second order.

Example A of the previous section provides a model in which the transition - at least according to the above mean field analysis - is second order (see Fig. 24). For $T \gg T_c$, the expectation value of the scalar field φ vanishes at all points $\underline{x}$ in space:

$$< \varphi(\underline{x}) >= 0 \,. \tag{5.15}$$

For $T < T_c$, this value of $< \varphi(\underline{x}) >$ becomes unstable and $< \varphi(\underline{x}) >$ evolves smoothly in time to a new value $\pm\sigma$. The direction is determined by thermal and quantum fluctuations and is therefore not uniform in space. There will be domains of average radius $\xi(t)$ in which $< \varphi(\underline{x}) >$ is coherent. By causality, the coherence length is bounded from above by the horizon. However, typical values of $\xi(t)$ are proportional to $\lambda^{-1}\sigma^{-1}$ if φ was in thermal equilibrium before the phase transition.

In condensed matter physics, a transition of the above type is said to proceed by spinodal decomposition[95], triggered by a rapid quench.

In Example B of the previous section, (see Fig. 25) the phase transition is first order. For $T > T_c$, the expectation value $< \varphi(x) >$ is approximately 0, the minimum of the high temperature effective potential. Provided the zero temperature potential has a sufficiently high barrier separating the metastable state $\varphi = 0$ from the global minimum (compared to the energy density in thermal fluctuations at $T = T_c$), then $\varphi(\underline{x})$ will remain trapped at $\varphi = 0$ also for $T < T_c$. In the notation of Ref. 96, the field φ is trapped in the false vacuum. After some time (determined again by the potential barrier), the false vacuum will decay by quantum tunnelling.

Tunnelling in quantum field theory was discussed in Refs. 96-99 (for reviews see e.g., Refs. 100 and 8). The transition proceeds by bubble nucleation. There is a probability per unit time and volume that at a point $\underline{x}$ in space a bubble of "true vacuum" $\varphi(\underline{x}) = a$ will nucleate. The nucleation radius is microscopical. As long as the potential barrier is large, the bubble radius will increase with the speed of light after nucleation. Thus, a bubble of $\varphi = a$ expands in a surrounding "sea" of false vacuum $\varphi = 0$.

To conclude, let us stress the most important differences between the two types of phase transitions discussed above. In a second order transition, the dynamics is determined mainly by classical physics. The transition occurs homogeneously in space (apart from the phase boundaries which – as discussed below – become topological defects), and $< \varphi(x) >$ evolves continuously in time. In first order transitions, quantum mechanics is essential. The process is extremely inhomogeneous, and $< \varphi(x) >$ is discontinuous as a function of time. As we shall see in the following sections, the above two types of transitions are the basis of various classes of inflationary Universe models.

5.4. Models of Inflation

At this stage we have established the formalism to be able to discuss models of inflation. I will focus on "old inflation," "new inflation"" and "chaotic inflation." There are many other attempts at producing an inflationary scenario, but there is as of now no convincing realization.

Old Inflation

The old inflationary Universe model[30,101] is based on a scalar field theory which undergoes a first order phase transition. As a toy model, consider a scalar field theory with the potential $V(\varphi)$ of Example B (see Fig. 25). Note that this potential is fairly general apart from the requirement that $V(a) = 0$, where $\varphi = a$ is the global minimum of $V(\varphi)$. This condition is required to avoid a large cosmological constant today (no inflationary Universe model manages to circumvent or solve the cosmological constant problem).

For fairly general initial conditions, $\varphi(x)$ is trapped in the metastable state $\varphi = 0$ as the Universe cools below the critical temperature T_c. As the Universe expands further, all contributions to the energy-momentum tensor $T_{\mu\nu}$ except for the contribution

$$T_{\mu\nu} \sim V(\varphi) g_{\mu\nu} \tag{5.16}$$

redshift. Hence, the equation of state approaches $p = -\rho$, and inflation sets in. Inflation lasts until the false vacuum decays. During inflation, the Hubble constant is given by

$$H^2 = \frac{8\pi G}{3} V(0)\,. \tag{5.17}$$

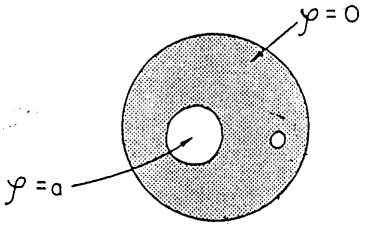

Figure 26: A sketch of the spatially inhomogeneous distribution of φ in Old Inflation.

After a period Γ^{-1}, where Γ is the tunnelling decay rate, bubbles of $\varphi = a$ begin to nucleate in a sea of false vacuum $\varphi = 0$. For a sketch of the resulting inhomogeneous distribution of $\varphi(x)$ see Fig. 26. Note that inflation stops after bubble nucleation.

The time evolution in old inflation is summarized in Fig. 27. We denote the beginning of inflation by t_i (here $t_i \simeq t_c$), the end by t_R (here $t_R \simeq t_c + \Gamma^{-1}$).

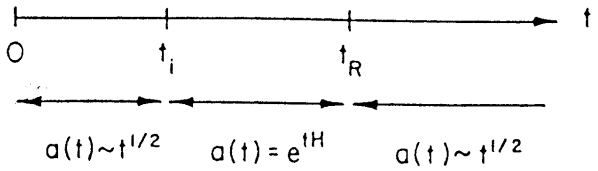

Figure 27: Phases in the old and new inflationary Universe.

It was immediately realized that old inflation has a serious "graceful exit" problem[102]. The bubbles nucleate after inflation with radius $r \ll 2t_R$ and would today be much smaller than our apparent horizon. Thus, unless bubbles percolate, the model predicts extremely large inhomogeneities inside the Hubble radius, in contradiction with the observed isotropy of the microwave background radiation.

For bubbles to percolate, a sufficiently large number must be produced so that they collide and homogenize over a scale larger than the present Hubble radius.

However, with exponential expansion, the volume between bubbles expands exponentially whereas the volume inside bubbles expands only with a low power. This prevents percolation.

New Inflation

Because of the graceful exit problem, old inflation never was considered to be a viable cosmological model. However, soon after the seminal paper by Guth, Linde and independently Albrecht and Steinhardt put forwards a modified scenario, the New Inflationary Universe[94] (see also Ref. 103).

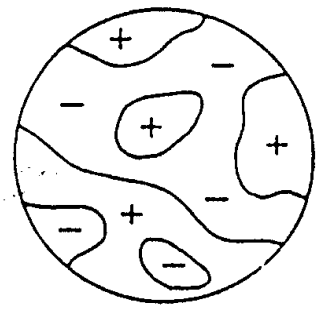

Figure 28: A sketch of the spatial distribution of φ in New Inflation after the transition. The symbols + and − indicate regions where $\varphi = +\sigma$ and $\varphi = -\sigma$ respectively.

The starting point is a scalar field theory with a double well potential which undergoes a second order phase transition (Fig. 24). $V(\varphi)$ is symmetric and $\varphi = 0$ is a local maximum of the zero temperature potential. Once again, it was argued that finite temperature effects confine $\varphi(\underline{x})$ to values near $\varphi = 0$ at temperatures $T \geq T_c$. For $T < T_c$, thermal fluctuations trigger the instability of $\varphi(\underline{x}) = 0$ and $\varphi(\underline{x})$ evolves towards $\varphi = \pm\sigma$ by the classical equation of motion

$$\ddot{\varphi} + 3H\dot{\varphi} - a^{-2} \bigtriangledown^2 \varphi = -V'(\varphi)\,. \tag{5.18}$$

The transition proceeds by spinodal decomposition (see Fig. 28) and hence

$\varphi(\underline{x})$ will be homogeneous within a correlation length. The analysis will be confined to such a small region. Hence, in Eq. (5.18) we can neglect the spatial gradient terms. Then, from (3.9) we can read off the induced equation of state. The condition for inflation is

$$\dot{\varphi}^2 \ll V(\varphi)\,, \tag{5.19}$$

i.e. slow rolling.

Often, the "slow rolling" approximation is made to find solutions of (5.18). This consists of dropping the $\ddot{\varphi}$ term. In this case, (5.18) becomes

$$3H\dot{\varphi} \;=-\,V'(\varphi)\,. \tag{5.20}$$

As an example, consider a potential which for $|\varphi| < \sigma$ has the following expansion near $\varphi = 0$

$$V(\varphi) = V_0 - \frac{1}{2}m^2\varphi^2\,. \tag{5.21}$$

With the above $V(\phi)$, (5.20) has the solution

$$\varphi(t) \;=\; \varphi(0)\exp\left(\frac{m^2}{3H}t\right) \tag{5.22}$$

(taking $H = \text{const}$ which is a good approximation). Thus, provided $m \ll \sqrt{3}H$, $\ddot{\varphi}$ is indeed smaller than the other terms in (5.18) and the slow rolling approximation seems to be satisfied.

However, the above conclusion is premature[104]. Equation (5.18) has a second solution. For $m > H$ the solution is

$$\varphi(t) \simeq \varphi(0)e^{mt} \tag{5.23}$$

and dominates over the previous one. This example shows that the slow rolling approximation must be used with caution. Here, however, the conclusion remains that provided $m \ll H$, then the model produces enough inflation to solve the cosmological problems.

There is no graceful exit problem in the new inflationary Universe. Since the spinodal decomposition domains are established before the onset of inflation, any boundary walls will be inflated outside the present Hubble radius.

The condition $m^2 \ll 3H^2$ which must be imposed in order to obtain inflation, is a fine tuning of the particle physics model – the first sign of problems with this scenario. Consider *e.g.* the model (5.11). By expanding $V(\varphi)$ about $\varphi = 0$ we can determine both H and m in terms of λ and σ. In order that $m^2 < 3H^2$ be satisfied we need

$$\sigma > \left(\frac{1}{6\pi}\right)^{1/2} m_{pl} \,, \tag{5.24}$$

which is certainly an unnatural constraint for models motivated by particle physics.

Let us, for the moment, return to the general features of the new inflationary Universe scenario. At the time t_c of the phase transition, $\varphi(t)$ will start to move from near $\varphi = 0$ towards either $\pm\sigma$ as described by the classical equation of motion, *i.e.* (5.22). At or soon after t_c, the energy-momentum tensor of the Universe will start to be dominated by $V(\varphi)$, and inflation will commence. t_i shall denote the time of the onset of inflation. Eventually, $\phi(t)$ will reach large values for which (5.21) is no longer a good approximation to $V(\varphi)$ and for which nonlinear effects become important. The time at which this occurs is t_B. For $t > t_B$, $\varphi(t)$ rapidly accelerates, reaches $\pm\sigma$, overshoots and starts oscillating about the global minimum of $V(\varphi)$. The amplitude of this oscillation is damped by the expansion of the Universe and (predominantly) by the coupling of φ to other fields. At time t_R, the energy in φ drops below the energy of the thermal bath of particles produced during the period of oscillation.

The evolution of $\varphi(t)$ is sketched in Fig. 29. The time period between t_B and t_R is called the reheating period and is usually short compared to the Hubble expansion time. The time evolution of the temperature T of the thermal radiation bath is also shown in Fig. 29.

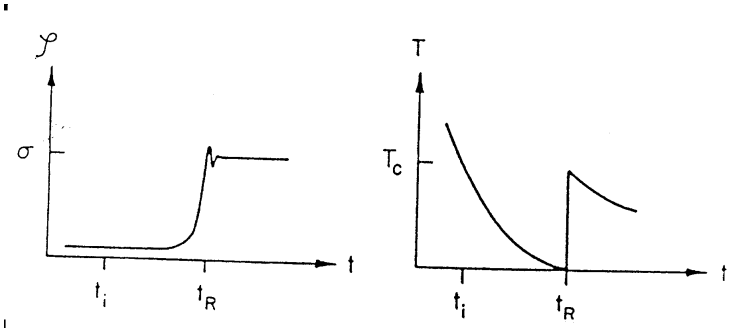

Figure 29: Evolution of $\varphi(t)$ and $T(t)$ in the new inflationary Universe.

Reheating in inflationary Universe models has been considered in Refs. 105-108. One way to view the process is as follows[107,108]. Consider a second scalar field ψ coupled to φ via the interaction Lagrangian

$$\mathcal{L}_I = \frac{1}{2} g \varphi^2 \psi^2 \,. \tag{5.25}$$

Then, an oscillating $\varphi(t)$ will act as a time dependent mass with periodic variations in the equation of motion for ψ

$$\ddot{\psi}_k + 3H\dot{\psi}_k + \left(m_\psi^2 + k^2 a^{-2}(t) + g\varphi^2(t)\right)\psi_k = 0 \tag{5.26}$$

where we have neglected nonlinear terms and expanded ψ into Fourier modes ψ_k. If the expansion of the Universe can be neglected and for periodic $\varphi^2(t)$, the above is the well known Mathieu equation[109] whose solutions have instabilities for certain values of k. These instabilities correspond to the production of ψ particles with well determined momenta[107,108]. These particles eventually equilibrate and regenerate a thermal bath.

For $t > t_R$, the Universe is again radiation dominated. Hence, the stages of the new inflationary Universe are the same as for old inflation (Fig. 27). There is a useful order of magnitude relation between the scale of symmetry breaking σ

and H. From

$$H^2 = \frac{8\pi G}{3} V(0) \tag{5.27}$$

and from the form of the potential (see (5.11)) it follows that

$$\left(\frac{H}{\sigma}\right) \sim \lambda^{1/2} \left(\frac{\sigma}{m_{pl}}\right) . \tag{5.28}$$

In particular, for $\sigma \sim 10^{15}$GeV (typical scale of grand unification) and $\lambda \sim 1$ we obtain $H \sim 10^{11}$GeV.

The new inflationary Universe model – although it was for a long time presented as a viable model – suffers from severe fine tuning and initial condition problems. In (5.24) we encountered the first of these problems: in order to obtain enough inflation, the potential must be fairly flat near $\varphi = 0$. A more severe problem will be derived in Section 5.5: Inflationary Universe models generate energy density perturbations. The steeper the potential, the larger the density perturbations. For a potential which near $\varphi = 0$ or $\varphi = H$ has the following expansion

$$V(\varphi) = V(0) - \lambda \varphi^4 , \tag{5.29}$$

the density perturbations conflict with observations unless (see later)

$$\lambda < 10^{-12} . \tag{5.30}$$

This in itself is an unexplained small number problem. However, even if we were willing to accept this we would run into initial condition problems[92,93]. For the new inflationary Universe to proceed in the way outlined above, it is essential that the field φ be in thermal equilibrium with other fields. This implies that the constant g coupling φ to other fields should not be too small. However, a coupling term of the form (5.25) induces one loop quantum corrections to the self coupling constant λ of the order g^2. Hence, the constraint $\lambda < 10^{-12}$ implies a constraint $g < 10^{-6}$. Thus, φ will not be in thermal equilibrium at t_c, and hence there will be no thermal forces which localize φ close to $\varphi = 0$.

Note that the above problem is not an artifact of using quartic potentials such as (5.29). Similar constraints would arise in other (*e.g.* quadratic) models. However, (5.29) was long considered to be the prototypical shape of $V(\varphi)$ for small values of φ since it is the shape which arises in Coleman-Weinberg[110] models.

In the absence of thermal forces which constrain φ to start close to $\varphi = 0$, the only constraints on $\varphi-$ at least using classical physics alone – come from energetic considerations. Obviously, it is unnatural to assume that at the initial time t_i the energy density in φ exceeds the energy density of one degree of freedom of the thermal bath at time t_i (temperature T_i). This implies

$$\begin{aligned} V(\varphi(\underline{x}, t_i)) &< \frac{\pi^2}{30} T_i^4 \\ |\bigtriangledown \varphi(\underline{x}, t_i)|^2 &< \frac{\pi^2}{30} T_i^4 a^2(t_i) \\ |\dot{\varphi}^2(\underline{x}, t_i)|^2 &< \frac{\pi^2}{30} T_i^4 \end{aligned} \tag{5.31}$$

In particular, for the double well potential of (5.11), (5.31) implies that $\varphi(\underline{x}, t_i)$ can be of the order

$$\varphi(\underline{x}, t_i) \sim \lambda^{-1/4} T_i \tag{5.32}$$

which for $T_i > \sigma$ is much larger than σ. In a weakly coupled model, the only natural time to impose initial conditions on $\varphi(\underline{x})$ is the Planck time, *i.e.* $T_i \sim m_{pl}$. Hence, the initial conditions allow and in fact suggest

$$\varphi(\underline{x}, t_i) \sim \lambda^{-1/4} m_{pl} \gg m_{pl} \quad \text{for } \mathrm{T_i} \sim \mathrm{T_{pl}}, \tag{5.33}$$

These observations lead to the chaotic inflation scenario, the only of the original inflationary Universe models which can still be considered as a viable scenario today.

Chaotic Inflation

Chaotic inflation[92] is based on the observation that for weakly coupled scalar fields, initial conditions which follow from classical considerations alone lead to very large values of $\varphi(\underline{x})$ (see (5.32)).

Consider a region in space where at the initial time $\varphi(\underline{x})$ is very large, homogeneous (we will make these assumptions quantitative below) and static. In this case, the energy-momentum tensor will be immediately dominated by the large potential energy term and induce an equation of state $p \simeq -\rho$ which leads to inflation. Due to the large Hubble damping term in the scalar field equation of motion, $\varphi(\underline{x})$ will only roll very slowly towards $\varphi = 0$. The kinetic energy contribution to $T_{\mu\nu}$ will remain small, the spatial gradient contribution will be exponentially suppressed due to the expansion of the Universe, and thus inflation persists. This is a brief survey of the chaotic inflation scenario. Note that in contrast to old and new inflation, no initial thermal bath is required. Note also that the precise form of $V(\varphi)$ is irrelevant to the mechanism. In particular, $V(\varphi)$ need not be a double well potential. This is a significant advantage, since for scalar fields other than Higgs fields used for spontaneous symmetry breaking, there is no particle physics motivation for assuming a double well potential, and since the inflaton (the field which gives rise to inflation) cannot be a conventional Higgs field due to the severe fine tuning constraints.

Let us consider the chaotic inflation scenario in more detail. For simplicity, take the potential

$$V(\varphi) = \frac{1}{2} m^2 \varphi^2 \tag{5.34}$$

and consider a region in space in which $\varphi(\underline{x}, t_i)$ is sufficiently homogeneous. To be specific, we require

$$\frac{1}{2} a^{-2}(t_i) | \bigtriangledown \varphi (\underline{x}, t_i)|^2 \ll V (\varphi(\underline{x}, t_i)) \tag{5.35}$$

over a region of size d_i

$$d_i \geq 3H^{-1}(t_i)\,. \tag{5.36}$$

We also require that the kinetic energy be negligible at the initial time t_i,

$$\dot{\varphi}(\underline{\mathrm{x}}, t_i)^2 \ll V(\varphi\,(\underline{\mathrm{x}}, t_i))\ , \tag{5.37}$$

although this assumption can be relaxed without changing the results[111]. From (5.35) and (5.37) it follows that at t_i the equation of state is inflationary, *i.e.* $p(t_i) \simeq -\rho(t_i)$. Condition (5.36) ensures that no large inhomogeneities can propagate from outside to the center of the region under consideration. With these approximations, the equation of motion for φ becomes

$$\ddot{\varphi} + 3H\dot{\varphi} = -m^2\varphi \tag{5.38}$$

with

$$H = \left(\frac{4\pi}{3}\right)^{1/2} \frac{m}{m_{pl}}\,\varphi\,. \tag{5.39}$$

Since we expect $\varphi(\underline{\mathrm{x}}, t)$ to be changing slowly, we make the slow rolling approximation

$$3H\dot{\varphi} = -m^2\varphi \tag{5.40}$$

which gives

$$\dot{\varphi} = -\left(\frac{1}{12\pi}\right)^{1/2} m\, m_{pl} \tag{5.41}$$

and shows that the approximation is self consistent. In order to get inflation, we

require

$$\frac{1}{2}\,\dot{\varphi}^2 < \frac{1}{2} m^2\,\varphi^2 \tag{5.42}$$

which (by (5.41)) is satisfied if

$$\varphi > \left(\frac{1}{12}\right)^{1/2} m_{pl}\,. \tag{5.43}$$

In order to obtain a period $\tau > 50 H^{-1}(t_i)$ of inflation, a slightly stronger condition is needed:

$$\varphi > 3 m_{pl}\,. \tag{5.44}$$

With chaotic inflation, the initial hope that grand unified theories could provide the answer to the homogeneity and flatness problems has been abandoned. The inflaton is introduced as a new scalar field (with no particular particle physics role) which is very weakly coupled to itself and to other fields (see Ref. 112 for an attempt to couple the inflaton to non-grand unified particle physics). In supergravity and in superstring inspired models there are scalar fields which are candidates to be the inflaton. I refer the reader to Refs. 113-115 for a discussion of this issue. However, even in such models the time when the inflaton φ decouples from the rest of physics is the Planck time $t_i = t_{pl}$. Thus, the chaotic inflation scenario is often called primordial inflation.

Chaotic inflation is a much more radical departure from standard cosmology than old and new inflation. In the latter, the inflationary phase can be viewed as a short phase of exponential expansion bounded at both ends by phases of radiation domination. In chaotic inflation, a piece of the Universe emerges with an inflationary equation of state immediately after the quantum gravity epoch.

The chaotic inflationary Universe scenario has been developed in great detail (see *e.g.*, Ref. 116 for a recent review). One important addition is the inclusion of stochastic noise[117] in the equation of motion for φ in order to take into account

the effects of quantum fluctuations. It can in fact be shown that for sufficiently large values of $|\varphi|$, the stochastic force terms are more important than the classical relaxation force $V'(\varphi)$. There is equal probability for the quantum fluctuations to lead to an increase or decrease of $|\varphi|$. Hence, in a substantial fraction of comoving volume, the field φ will climb up the potential. This leads to the conclusion that chaotic inflation is eternal. At all times, a large fraction of the physical space will be inflating. Another consequence of including stochastic terms is that on large scales (much larger than the present Hubble radius), the Universe will look extremely inhomogeneous.

General Comments

Old, new and choatic inflation are all based on the use of new fundamental scalar fields which cannot be the Higgs field of an ordinary gauge theory. Instead of introducing new physics via scalar fields – an approach which makes the cosmological constant problem more severe – it is possible to look for realizations of inflation based on some alternative new physics which do not invoke fundamental scalar fields.

One possibility is to consider modifications of Einstein gravity which can lead to inflation. In fact, the first model of inflation[39] was based on considering an action for gravity of the form

$$S = \int d^4x \sqrt{-g}\,(R + \varepsilon R^2)\,, \tag{5.45}$$

where R is the Ricci scalar of the metric $g_{\mu\nu}$. As shown first by Whitt[118], the equations of motion resulting from this action are the same as those from Einstein gravity in the presence of a scalar field with a special potential which allows for chaotic inflation. The relationship is obtained via a conformal transformation.

Since perturbative quantum gravity calculations and considerations based on quantum field theory in curved space-time all point to the presence of higher derivative terms in the action for gravity, it is not unlikely that successful inflation will

come from the gravity sector. A recently proposed theory[41], in which the curvature is bounded for all solutions, predicts that our Universe will have started out in an inflationary period.

Another interesting possibility is that inflation is a result of new physics associated with a unified theory of all forces such as superstring theory. Interesting speculations along these lines have recently been made in Ref. 119.

As has hopefully become clear, inflation is a nice idea which solves many problems of standard big bang cosmology. However, no convincing realization of inflation which does not involve unexplained small numbers has emerged (for a general discussion of this point see Ref. 120).

It is important to distinguish between models of inflation which are self consistent and those which are not. We have shown that new inflation is not self consistent, whereas chaotic inflation is. One of the key issues involves initial conditions. In new inflation, the initial conditions required can only be obtained if the inflaton field is in thermal equilibrium above the critical temperature, which however is not possible because of the density fluctuation constraints on coupling constants.

In chaotic inflation, it can be shown that – provided the spatial sections are flat – a large phase space of initial conditions (much larger than is apparent from (5.35) and (5.37)) gives chaotic inflation[121–123], whereas the probability to relax dynamically[124] to field configurations which give new inflation (this possibility is only available in double well potentials) is negligibly small.

5.5. Generation and Evolution of Fluctuations

Preliminaries

In this section, the origin of the primordial density perturbations required to seed galaxies will be discussed within the context of inflationary Universe models. From Chapter 3, we recall the basic reason why in inflationary cosmology a causal

generation mechanism is possible: comoving scales of cosmological interest today originate inside the Hubble radius early in the de Sitter period (see Fig. 11). Hence, it is in principle possible that density perturbations on these scales can be generated by a causal mechanism at very early times.

First, let us demonstrate why the Hubble radius $H^{-1}(t)$ is the length scale of relevance in these considerations. Consider a scalar field theory with action

$$S(\varphi) = \int d^4x\sqrt{-g}\,\left[\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V(\varphi)\right]\,. \tag{5.46}$$

The resulting equation of motion is

$$\ddot{\varphi} + 3H\dot{\varphi} - a^{-2}\nabla^2\varphi = -V'(\varphi)\,. \tag{5.47}$$

The second term on the left hand side is the Hubble damping term, the third represents microphysics (spatial gradients). To simplify the consideration, assume that $V'(\varphi) = 0$. Then, the time evolution is influenced by microphysics and gravity. For plane wave perturbations with wave number k, the gravitational force is proportional to $H^2\varphi$ whereas the microphysical force is $a^{-2}k^2\varphi$. Thus, for $ak^{-1} < H^{-1}$, i.e., on length scales smaller than the Hubble radius, microphysics dominates, whereas for $ak^{-1} > H^{-1}$, i.e., on length scales larger than the Hubble radius, the gravitational drag dominates.

Based on the above analysis we can formulate the main idea of the fluctuation analysis in inflationary cosmology. In linear order, all Fourier modes decouple. Hence, we fix a mode with wave number k. There are two very different time intervals to consider. Let $t_i(k)$ be the time when the scale crosses the Hubble radius in the de Sitter phase, and $t_f(k)$ the time when it reenters the Hubble radius after inflation (see Fig. 30). The first period lasts until $t = t_i(k)$. In this time interval microphysics dominates. We shall demonstrate that quantum fluctuations generate perturbations during this period[35–38,125]. The second time interval is $t_i(k) < t < t_f(k)$. Now microphysics is unimportant and the evolution of perturbations is

determined by gravity. At $t_i(k)$, decoherence sets in; the quantum mechanical wave functional evolves as a statistical ensemble of classical configurations after this time.

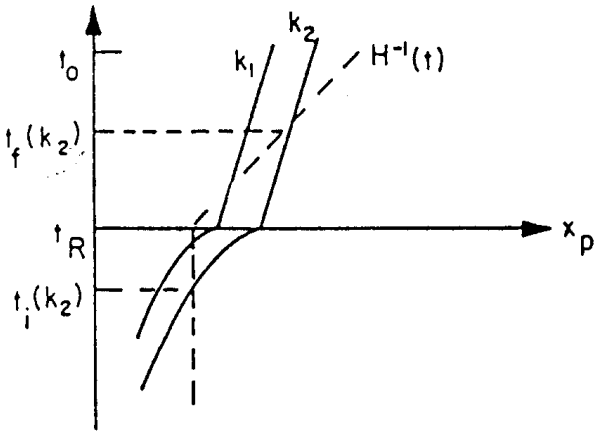

Figure 30: Sketch (physical distance x_p versus time t) of the evolution of two fixed comoving scales labelled k_1 and k_2 in the inflationary Universe.

It is possible to give a heuristic derivation of the shape of the spectrum of cosmological perturbations in an inflationary Universe based on simple geometrical arguments[35]. Consider first the process of generation of fluctuations. If the generation mechanism produces inhomogeneities at all times on a fixed physical wavelength with an amplitude determined by the "Hawking temperature" T_H of de Sitter space[126],

$$T_H = \frac{H}{2\pi}\,, \tag{5.48}$$

then the evolution of perturbations on different scales between when they are formed and when they leave the Hubble radius at times $t_i(k)$ is related by time translation, and the amplitude of the fluctuations when measured at time $t_i(k)$ will be independent of k:

$$\frac{\delta M}{M}(k, t_i(k)) = \mathcal{A}_i = \text{ const.} \tag{5.49}$$

In fact, the amplitude $\mathcal{A}_i$ will be given by the thermal energy associated with

the Hawking temperature (5.49) divided by the "false vacuum" energy density ρ_0 responsible for inflation

$$\mathcal{A}_i \sim \frac{T_H^4}{\rho_0} \,. \tag{5.50}$$

In Chapter 4 it was shown (see (4.68)) that the gauge invariant measure of density fluctuations evolves trivially on scales outside of the Hubble radius: it changes by a factor which only depends on the equations of state in the inflationary and past-inflationary phases. In particular, this factor is independent of k. Hence, using (4.68), we arrive at the conclusion that perturbations are independent of scale when they enter the Hubble radius

$$\frac{\delta M}{M}(k, t_f(k)) = \mathcal{A}_f = \text{ const} \,. \tag{5.51}$$

Hence, we have demonstrated that a scale invariant spectrum of density perturbations is a generic feature of an inflationary Universe scenario.

From (4.67) and (4.68) if follows that

$$\mathcal{A}_f \sim \mathcal{A}_i \, \frac{1 + w(t_f(k))}{1 + w(t_i(k))} \,. \tag{5.52}$$

On scales which enter the Hubble radius after t_{eq},

$$1 + w(t_f(k)) = 1 \,. \tag{5.53}$$

The initial value of $w = p/\rho$ during inflation can be determined by using the expressions (3.9) for energy density and pressure of the scalar field responsible for inflation. The result is

$$1 + w(t_i(k)) = \frac{\dot{\varphi}^2(t_i(k))}{\rho(t_i(k))} \sim \frac{H^4}{\rho_0} \,, \tag{5.54}$$

where the last proportionality follows by dimensional analysis. Thus, combining

(5.50-5.54) we obtain

$$\frac{\delta M}{M}(k, t_f(k)) \sim 1\,, \tag{5.55}$$

which is at least four orders of magnitude too large to conform with the constraints from CMB anisotropy measurements.

The above problem is known as the "fluctuation problem"[127,90] and is common to most inflationary Universe models. The only known solutions involve small numbers introduced into the particle physics sector. This defeats one of the aims of inflation which is to avoid the need for unnaturally small constants. To study this problem in more detail we must turn to a quantitative analysis of the generation and evolution of fluctuations.

Quantum Generation of Fluctuations

The question of the origin of classical density perturbations from quantum fluctuations in the de Sitter phase of an inflationary Universe is a rather subtle issue. Starting from a homogeneous quantum state (*e.g.*, the vacuum state in the FRW coordinate frame at time t_i, the beginning of inflation), a naive semiclassical anaylsis would predict the absence of fluctuations since

$$< \psi | T_{\mu\nu}(x) | \psi >= \text{const}\,. \tag{5.56}$$

However, as a simple thought experiment shows, such a naive analysis is inappropriate. Imagine a local gravitational mass detector D positioned close to a large mass M which is suspended from a pole (see Fig. 31). The decay of an alpha particle will sever the cord (at point T) by which the mass is held to the pole and the mass will drop. According to the semiclassical prescription

$$G_{\mu\nu} = 8\pi G < \psi | T_{\mu\nu} | \psi >\,, \tag{5.57}$$

the metric (*i.e.*, the mass measured) will slowly decrease. This is obviously not what happens. The mass detector shows a signal which corresponds to one of the

classical trajectories which make up the state $|\psi>$, a trajectory corresponding to a sudden drop in the gravitational force measured.

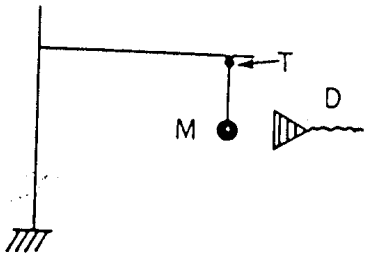

Figure 31: Sketch of the thought experiment discussed in the text.

The origin of classical density perturbations as a consequence of quantum fluctuations in a homogeneous state $|\psi>$ can be analyzed along similar lines. The quantum to classical transition is picking out[128−130] one of the typical classical trajectories which make up the wave function of $|\psi>$. We can implement[131,132] the procedure as follows: Define a classical scalar field

$$\varphi_{cl}(\underline{x}, t) = \varphi_0(t) + \delta\varphi(\underline{x}, t) \tag{5.58}$$

with vanishing spatial average of $\delta\varphi$. The induced classical energy momentum tensor $T^{cl}_{\mu\nu}(\underline{x}, t)$ which is the source for the metric is given by

$$T^{cl}_{\mu\nu}(\underline{x}, t) = T_{\mu\nu}(\varphi_{cl}(\underline{x}, t))\,, \tag{5.59}$$

where $T_{\mu\nu}\left(\varphi_{cl}(\underline{x}, t)\right)$ is defined as the canonical energy-momentum tensor of the classical scalar field $\varphi_{cl}(\underline{x}, t)$. Unless $\delta\varphi$ vanishes, $T^{cl}_{\mu\nu}$ is inhomogeneous.

For applications to chaotic inflation, we take $|\psi>$ to be a Gaussian state with mean value $\varphi_0(t)$

$$<\psi|\varphi^2(\underline{x}, t)|\psi> = \varphi_0^2(t)\,. \tag{5.60}$$

Its width is taken to be the width of the vacuum state of the free scalar field theory with mass determined by the curvature of $V(\varphi)$ at φ_0. This state is used to define

the Fourier transform $\delta\tilde{\varphi}(k,t)$ by

$$|\delta\tilde{\varphi}(k)|^2 =< \psi| \, |\tilde{\varphi}(k)|^2 \, |\psi > \ . \tag{5.61}$$

The amplitude of $\delta\tilde{\varphi}(k)$ is identified with the width of the ground state wave function of the harmonic osciallator $\tilde{\varphi}(k)$. (Recall that each Fourier mode of a free scalar field is a harmonic oscillator). Note that no divergences arise in the above construction. In principle, quantum fluctuations contribute a term to $\varphi_0(t)$; this backreaction effect has not yet been studied. The quantum corrections to (5.60) are divergent and must be regularized and renormalized (see *e.g.*, Ref. 133). They are the source of the stochastic driving forces in stochastic chaotic inflation.

By linearizing (5.59) about $\varphi_0(t)$ we obtain the perturbation of the energy-momentum tensor. In particular, the energy density fluctuation $\delta\tilde{\rho}(k)$ is given by

$$\delta\tilde{\rho}(k) = \dot{\varphi}_0 \delta\dot{\tilde{\varphi}}(k) + V'(\varphi_0)\delta\tilde{\varphi}(k) \, . \tag{5.62}$$

To obtain the initial amplitude $\mathcal{A}_i$ of (5.49), the above is to be evaluated at the time $t_i(k)$.

The computation of the spectrum of density perturbations produced in the de Sitter phase has been reduced to the evaluation of the expectation value (5.61). First, we must specify the state $|\psi >$. (Recall that in non-Minkowski space-times there is no uniquely defined vacuum state of a quantum field theory). We pick the FRW frame of the pre-inflationary period. In this frame, the number density of particles decreases exponentially. Hence we choose $|\psi >$ to be the ground state in this frame (see Ref. 134 for a discussion of other choices). $\psi[\tilde{\varphi}(\underline{\mathrm{k}}), t]$, the wave functional of $|\psi >$, can be calculated explicitly. It is basically the superposition of the ground state wave functions for all oscillators

$$\psi[\tilde{\varphi}(\underline{\mathrm{k}}), t] = N \exp\left\{ -\frac{1}{2}(2\pi)^{-3} a^3(t) \int d^3k \omega(\underline{\mathrm{k}}, t) |\tilde{\varphi}(\underline{\mathrm{k}})|^2 \right\} \, . \tag{5.63}$$

N is a normalization constant and $\omega(\underline{\mathrm{k}}, t) \sim H$ at $t = t_i(k)$. Hence

$$\delta\tilde{\varphi}(\underline{\mathrm{k}}, t) = (2\pi)^{3/2} a^{-3/2} \omega(\underline{\mathrm{k}}, t)^{-1/2} \sim (2\pi)^{3/2} k^{-3/2} H \, , \; t = t_i(k) \, . \tag{5.64}$$

Evolution of Fluctuations

Given the above determination of the intitial amplitude of density perturbations at the time when they leave the Hubble radius during the de Sitter phase, and the general relativistic analysis of the evolution of fluctuations discussed in Section 4.4, it is easy to evaluate the r.m.s. inhomogeneities when they reenter the Hubble radius at time $t_f(k)$.

First, we combine (5.62), (5.64), (5.49), (4.10) and (4.15) to obtain

$$\left(\frac{\delta M}{M}\right)^2 (k, t_i(k)) \sim k^3 \left(\frac{V'(\varphi_0)\delta\tilde{\varphi}(k, t_i(k))}{\rho_0}\right)^2 \sim \left(\frac{V'(\varphi_0)H}{\rho_0}\right)^2 , \tag{5.65}$$

and thus

$$\mathcal{A}_i \sim \frac{V'(\varphi_0(t_i(k))H}{\rho_0} \, . \tag{5.66}$$

If the background scalar field is rolling slowly, then

$$V'(\varphi_0(t_i(k))) = 3H|\dot{\varphi}_0(t_i(k))| \, . \tag{5.67}$$

Combining (5.66) and (5.67) with (5.52) and (5.54) we get

$$\frac{\delta M}{M}(k, \, t_f(k)) = \mathcal{A}_f \sim \frac{3H^2|\dot{\varphi}_0(t_i(k))|}{\dot{\varphi}_0^2(t_i(k))} = \frac{3H^2}{|\dot{\varphi}_0(t_i(k))} = \frac{3H^2}{|\dot{\varphi}_0(t_i(k))|} \tag{5.68}$$

This result can now be evaluated for specific models of inflation to find the conditions on the particle physics parameters which give a value

$$\mathcal{A}_f \sim 10^{-5} \tag{5.69}$$

which is required if quantum fluctuations from inflation are to provide the seeds for galaxy formation and agree with the CMB anisotropy limits.

For chaotic inflation with a potential

$$V(\varphi) = \frac{1}{2} m^2 \varphi^2 \,, \tag{5.70}$$

the dynamics of φ was analyzed in Section 5b (see in particular (5.43) and (5.44)). We have

$$\varphi(t_i(k)) \sim m_{pl} \tag{5.71}$$

and hence

$$H(t_i(k)) \sim m_{pl}^{-1} m \, \varphi(t_i(k)) \sim m \,. \tag{5.72}$$

Therefore,

$$\frac{\delta M}{M}(k, t_f(k)) \sim 3 \frac{H^2}{|\dot{\varphi}_0(t_i(k))|} \sim 10 \frac{m}{m_{pl}} \tag{5.73}$$

which implies that

$$m \sim 10^{13} \, \mathrm{GeV} \tag{5.74}$$

to agree with (5.70).

Similarly, for a potential of the form

$$V(\varphi) = \frac{1}{4} \lambda \varphi^4 \tag{5.75}$$

we obtain

$$\frac{\delta M}{M}(k, \, t_f(k)) \sim 10 \cdot \lambda^{1/2} \tag{5.76}$$

which requires

$$\lambda \leq 10^{-12}. \tag{5.77}$$

in order not to conflict with observations.

The conditions (5.74) and (5.77) require the presence of small parameters in the particle physics model. It has been shown quite generally[120)] that small parameters are required if inflation is to solve the fluctuation problem.

Discussion

Let us first summarize the main results of the analysis of density fluctuations in inflationary cosmology:

- Quantum vacuum fluctuations in the de Sitter phase of an inflationary Universe are the source of perturbations.
- The quantum perturbations decohere on scales outside the Hubble radius and can hence be treated classically.
- The classical evolution outside the Hubble radius produces a large amplification of the perturbations. In fact, unless the particle physics model contains very small coupling constants, the predicted fluctuations are in excess of those allowed by the bounds on cosmic microwave anisotropies.
- Inflationary Universe models generically produce a scale invariant Harrison-Zel'dovich spectrum

$$\frac{\delta M}{M}(k, t_f(k)) = \text{const.} \tag{5.78}$$

It is not hard to construct models which give a different spectrum. All that is required is a significant change in H during the period of inflation.

I have chosen to present the analysis of fluctuations in inflationary cosmology in two separate steps in order to highlight the crucial physics issues. Having done this, it is possible to step back and construct a unified analysis in which expectation values of gauge invariant variables are propagated from $t \ll t_i(k)$ to $t_f(k)$ in a consistent way[135,11], and in which the final values of the expectation values of quadratic operators are used to construct $T^{cl}_{\mu\nu}(\underline{x}, t)$.

Once inside the Hubble radius, the evolution of the mass perturbations is influenced by the damping effects discussed in Section 4.3, which is turn depend on the composition of the dark matter. The dominant effects are the Meszaros effect and free streaming.

On scales which enter the Hubble radius before t_{eq}, the perturbations can only grow logarithmically in time between $t_f(k)$ and t_{eq}. This implies that (up to logarithmic corrections), the mass perturbation spectrum is flat for wavelengths smaller than λ_{eq}, the comoving Hubble radius at t_{eq}:

$$\frac{\delta M}{M}(\lambda, t) \simeq \text{const}, \quad t \leq t_{eq},\ \lambda < \lambda_{eq}, \tag{5.79}$$

whereas on larger scales

$$\frac{\delta M}{M}(\lambda, t) \propto \lambda^{-2}. \tag{5.80}$$

Equations (5.79) and (5.80) give the power spectrum in an $\Omega = 1$ inflationary CDM model.

If the dark matter is hot, then neutrino free streaming cuts off the power spectrum at λ_J^{max} (see (3.27)). The inflationary HDM and CDM perturbation spectra are compared in Fig. 32.

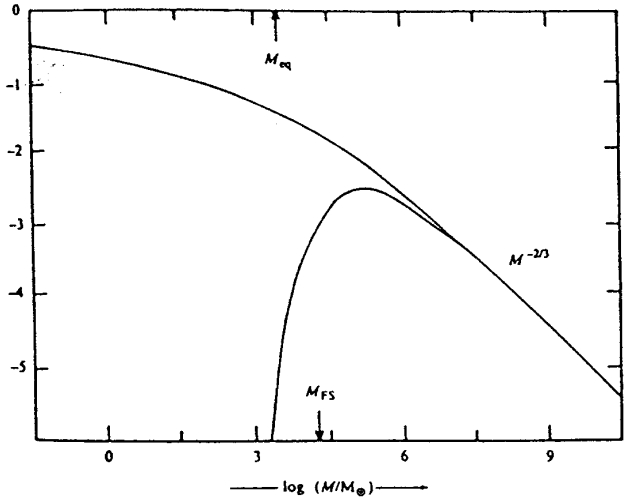

Figure 32: The linear theory power spectra for inflationary CDM (upper curve) and HDM models. The horizontal axis is mass, expressed in units of solar masses. M_{eq} is the mass inside the comoving horizon at t_{eq}, and M_{FS} is the mass inside the maximal comoving neutrino free streaming volume.

6. Topological Defects and Structure Formation

6.1. Classification

In the previous section we have seen that symmetry breaking phase transitions in unified field theories arising in particle physics (*e.g.,* Grand Unified Theories (GUT)[46]) do not lead, in general, to inflation. In most models, the coupling constants which arise in the effective potential for the scalar field φ driving the phase transition are too large to generate a period of slow rolling which lasts more than one Hubble time $H^{-1}(t)$. Nevertheless, there are interesting remnants of the phase transition: topological defects.

Consider a single component real scalar field with a typical symmetry breaking potential

$$V(\varphi) = \frac{1}{4}\lambda(\varphi^2 - \eta^2)^2 \tag{6.1}$$

Unless $\lambda \ll 1$ there will be no inflation. The phase transition will take place on a short time scale $\tau < H^{-1}$, and will lead to correlation regions of radius $\xi < t$ inside of which φ is approximately constant, but outside of which φ ranges randomly over the vacuum manifold $\mathcal{M}$, the set of values of φ which minimizes $V(\varphi)$ – in our example $\varphi = \pm\eta$. The correlation regions are separated by domain walls, regions in space where φ leaves the vacuum manifold $\mathcal{M}$ and where, therefore, potential energy is localized. Via the usual gravitational force, this energy density can act as a seed for structure.

Topological defects are familiar from solid state and condensed matter systems. Crystal defects, for example, form when water freezes or when a metal crystallizes[136]. Point defects, line defects and planar defects are possible. Defects are also common in liquid crystals. They arise in a temperature quench from the disordered to the ordered phase[137]. Vortices in ^{4}He are analogs of global cosmic strings. Vortices and other defects are also produced during a quench below the critical temperature in ^{3}He[138]. Finally, vortex lines also play an important role in the theory of superconductivity[139].

The analogies between defects in particle physics and condensed matter physics are quite deep. Defects form for the same reason: the vacuum manifold is topologically nontrivial. The arguments which say that in a theory which admits defects, such defects will inevitably form, are applicable both in cosmology and in condensed matter physics. Different, however, is the defect dynamics. The motion of defects in condensed matter systems is friction-dominated, whereas the defects in cosmology obey relativistic equations, second order in time derivatives, since they come from a relativistic field theory.

After these general comments we turn to a classification of topological defects. We consider theories with an n-component order parameter φ and with a potential energy function (free energy density) of the form (6.1) with

$$\varphi^2 = \sum_{i=1}^{n} \varphi_i^2 \,. \tag{6.2}$$

There are various types of local and global topological defects (regions of trapped energy density) depending on the number n of components of φ. The more rigorous mathematical definition refers to the homotopy of $\mathcal{M}$. The words "local" and "global" refer to whether the symmetry which is broken is a gauge or global symmetry. In the case of local symmetries, the topological defects have a well defined core outside of which φ contains no energy density in spite of nonvanishing gradients $\nabla\varphi$: the gauge fields A_μ can absorb the gradient, *i.e.,* $D_\mu\varphi = 0$ when $\partial_\mu\varphi \neq 0$, where the covariant derivative D_μ is defined by

$$D_\mu = \partial_\mu + ie\, A_\mu \,, \tag{6.3}$$

e being the gauge coupling constant. Global topological defects, however, have long range density fields and forces.

Table 1 contains a list of topological defects with their topological characteristic. A "v" markes acceptable theories, a "x" theories which are in conflict with observations (for $\eta \sim 10^{16}$ GeV).

Table 1	n	topology	local defect	global defect
domain wall	1	$\Pi_0(\mathcal{M}) \neq 1$	x	x
cosmic string	2	$\Pi_1(\mathcal{M}) \neq 1$	v	v
monopole	3	$\Pi_2(\mathcal{M}) \neq 1$	x	v
texture	4	$\Pi_3(\mathcal{M}) \neq 1$	-	v

Theories with domain walls are ruled out[140] since a single domain wall stretching across the Universe today would overclose the Universe. Local monopoles are also ruled out[141] since they would overclose the Universe. Local textures are ineffective at producing structures (see Section 6).

We now describe examples of domain walls, cosmic strings, monopoles and textures, focussing on configurations with maximal symmetry.

Domain walls

arise in theories with a single real order parameter and free energy density given by (6.1). The vacuum manifold of this model consists of two points

$$\mathcal{M} = \{\pm\eta\} \tag{6.4}$$

and hence has nontrivial zeroth homotopy group:

$$\Pi_0(\mathcal{M}) \neq 1 \tag{6.5}$$

(readers not familiar with homotopy groups can simply skip all of the following statements involving $\Pi_n(\mathcal{M})$. They are not required for an understanding of the physics).

To construct a domain wall configuration with planar symmetry (without loss of generality the $y-z$ plane can be taken to be the plane of symmetry), assume

that

$$\begin{aligned} \varphi(x) &\simeq \eta \quad x \gg \eta^{-1} \\ \varphi(x) &\simeq -\eta \quad x \ll -\eta^{-1} \end{aligned} \tag{6.6}$$

By continuity of φ, there must be an intermediate value of x with $\varphi(x) = 0$. We can take this point to be $x = 0$, *i.e.*,

$$\varphi(0) = 0\,. \tag{6.7}$$

The set of points with $\varphi = 0$ constitute the center of the domain wall. Physically, the wall is a thin sheet of trapped energy density. The width w of the sheet is given by the balance of potential energy and tension energy. Assuming that the spatial gradients are spread out over the thickness w we obtain

$$wV(0) = w\lambda\eta^4 \sim \frac{1}{w}\,\eta^2 \tag{6.8}$$

and thus

$$w \sim \lambda^{-1/2}\eta^{-1}\,. \tag{6.9}$$

See Fig. 33 for a sketch of this domain wall configuration.

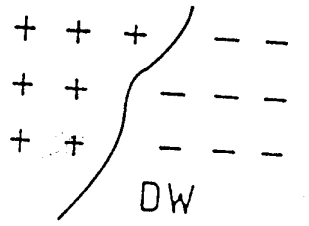

Figure 33: A low temperature field configuration containing a domain wall. The sketch shows a plane P in space. At positions with a $+$ or $-$, the value of the scalar field is η or $-\eta$, respectively. The solid line is the intersection with P of the plane of points in space with $\varphi = 0$.

A theory with a complex order parameter ($n = 2$) admits cosmic strings. In this case the vacuum manifold of the model is

$$\mathcal{M} = S^1 \,, \tag{6.10}$$

which has nonvanishing first homotopy group:

$$\Pi_1(\mathcal{M}) = Z \neq 1 \,. \tag{6.11}$$

A cosmic string is a line of trapped energy density which arises whenever the field $\varphi(x)$ circles $\mathcal{M}$ along a closed path in space (*e.g.*, along a circle). In this case, continuity of φ implies that there must be a point with $\varphi = 0$ on any sheet bounded by the closed path. The points on different sheets connect up to form a line overdensity of field energy (see Fig. 34).

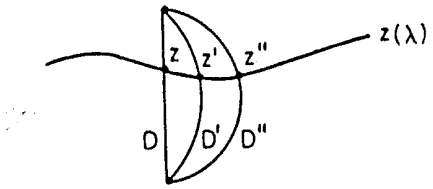

Figure 34: Sketch of the topological argument for the existence of cosmic string configurations. Given a field configuration with nontrivial winding along a circle normal to the plane of this figure, there must be a point with $\varphi = 0$ on every disk bounded by the circle. Three disks are depicted: D, D' and D'', and the respective points with $\varphi = 0$ are z, z' and z''. The union of all such points z forms the center $z(\lambda)$ of the string.

To construct a field configuration with a string along the z axis[43)], take $\varphi(x)$ to cover $\mathcal{M}$ along a circle with radius r about the point $(x, y) = (0, 0)$:

$$\varphi(r, \vartheta) \simeq \eta e^{i\vartheta} \,,\; r \gg \eta^{-1} \,. \tag{6.12}$$

This configuration has winding number 1, *i.e.*, it covers $\mathcal{M}$ exactly once. Main-

taining cylindrical symmetry, we can extend (6.12) to arbitrary r

$$\varphi(r,\,\vartheta) = f(r)e^{i\vartheta}\,, \tag{6.13}$$

where $f(0) = 0$ and $f(r)$ tends to η for large r. The width w can again be found by balancing potential and tension energy. The result is identical to the result (6.9) for domain walls.

For local cosmic strings, *i.e.*, strings arising due to the spontaneous breaking of a gauge symmetry, the energy density decays exponentially for $r \gg \eta^{-1}$. In this case, the energy μ per unit length of a string is finite and depends only on the symmetry breaking scale η

$$\mu \sim \eta^2 \tag{6.14}$$

(independent of the coupling λ). The value of μ is the only free parameter in a cosmic string model.

To see how the finiteness of the mass per unit length μ comes about, consider the simplest theory admitting local strings, the Abelian Higgs model with Lagrangean

$$\mathcal{L} = \frac{1}{2}\,D_\mu\varphi D^\mu\varphi - V(\varphi) + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{6.15}$$

where φ is a complex order parameter with potential (6.1), D_μ is the gauge covariant derivative

$$D_\mu = \partial_\mu + ie\,A_\mu\,, \tag{6.16}$$

the field A_μ is a U(1) gauge potential with associated field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\,, \tag{6.17}$$

and e is the gauge coupling constant.

For an order parameter configuration (6.12), the gauge fields A_μ will take on values such that

$$D_\mu \varphi \simeq 0 \quad r \gg \eta^{-1} \tag{6.18}$$

even though $\partial_\mu \varphi \neq 0$. Hence, the energy density decays exponentially for $r \gg \eta^{-1}$. For strings in a global theory (no gauge potential), the spatial gradient energy $(\partial_\mu \varphi)^2$ cannot be cancelled at large r, and hence the mass per unit length is logarithmically divergent as a function of a large r cutoff.

If the order parameter of the model has three components ($n = 3$), then monopoles result as topological defects. The vacuum manifold is

$$\mathcal{M} = S^2 \tag{6.19}$$

and has topology given by

$$\Pi_2(\mathcal{M}) \neq 1 \,. \tag{6.20}$$

Given a sphere S is space, it is possible that φ takes on values in $\mathcal{M}$ everywhere on S, and that it covers $\mathcal{M}$ once. By continuity, there must be a point in space in the interior of S with $\varphi = 0$. This is the center of a point-like defect, the monopole.

To construct a spherically symmetric monopole with the origin as its center, consider a field configuration φ which defines a map from physical space to field space such that all spheres S_r in space of radius $r \gg \eta^{-1}$ about the origin are mapped onto $\mathcal{M}$ (see Fig. 35):

$$\begin{aligned} \varphi : S_r &\longrightarrow \mathcal{M} \\ (r, \vartheta, \varphi) &\longrightarrow (\vartheta, \varphi) \,. \end{aligned} \tag{6.21}$$

This configuration defines a winding number one magnitude monopole.

Domain walls, cosmic strings and monopoles are examples of topological defects. A topological defect has a well-defined core, a region in space with $\varphi \notin \mathcal{M}$ and

hence $V(\varphi) > 0$. There is an associated winding number which is quantized, *i.e.*, it can take on only integer values. Since the winding number can only change continuously, it must be conserved, and hence topological defects are stable. Furthermore, topological defects exist for theories with global and local symmetries.

Now, let us consider a theory with a four-component order parameter (*i.e.*, $n = 4$), and a potential given by (6.1). In this case, the vacuum manifold is

$$\mathcal{M} = S^3 \tag{6.22}$$

and the associated topology is given by

$$\Pi_3(\mathcal{M}) \neq 1\,. \tag{6.23}$$

The corresponding defects are called "textures".[44,45]

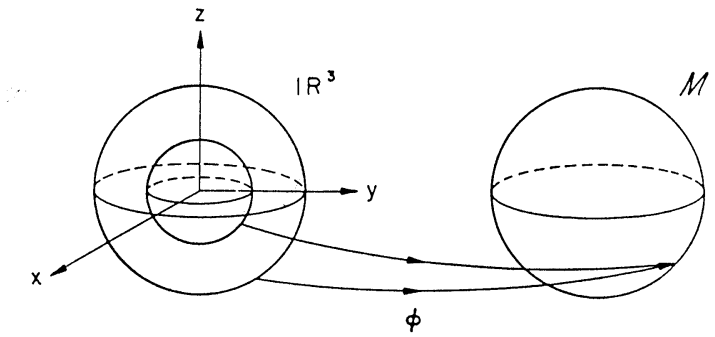

Figure 35: Construction of a monopole: left is physical space, right the vacuum manifold. The field configuration ϕ maps spheres in space onto $\mathcal{M}$. However, a core region of space near the origin is mapped onto field values not in $\mathcal{M}$.

Textures, however, are quite different than the previous topological defects. The texture construction will render this manifest (Fig. 36). To construct a radially symmetric texture, we give a field configuration $\varphi(x)$ which maps physical space onto $\mathcal{M}$. The origin 0 in space (an arbitrary point which will be the center

of the texture) is mapped onto the north pole N of $\mathcal{M}$. Spheres surrounding 0 are mapped onto spheres surrounding N. In particular, some sphere with radius $r_c(t)$ is mapped onto the equator sphere of $\mathcal{M}$. The distance $r_c(t)$ can be defined as the radius of the texture. Inside this sphere, $\varphi(x)$ covers half the vacuum manifold. Finally, the sphere at infinity is mapped onto the south pole of $\mathcal{M}$. The configuration $\varphi(\underline{\mathrm{x}})$ can be parameterized by[60)]

$$\varphi(x,y,z) = \left(\cos\chi(r),\ \sin\chi(r)\frac{x}{r},\ \sin\chi(r)\frac{y}{r},\ \sin\chi(r)\frac{z}{r}\right) \tag{6.24}$$

in terms of a function $\chi(r)$ with $\chi(0) = 0$ and $\chi(\infty) = \pi$. Note that at all points in space, $\varphi(\underline{\mathrm{x}})$ lies in $\mathcal{M}$. There is no defect core. All the energy is in spatial gradient (and possibly kinetic) terms.

In a cosmological context, there is infinite energy available in an infinite space. Hence, it is not necessary that $\chi(r) \to \pi$ as $r \to \infty$. We can have

$$\chi(r) \to \chi_{\max} < \pi \ \text{ as } \ r \to \infty\,. \tag{6.25}$$

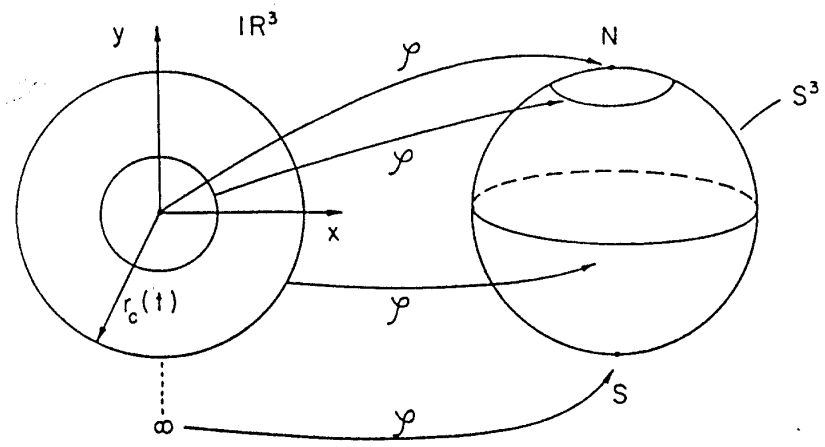

Figure 36:
Construction of a global texture: left is physical space, right the vacuum manifold. The field configuration ϕ is a map from space to the vacuum manifold (see text).

In this case, only a fraction

$$n_W = \frac{\chi_{\max}}{\pi} - \frac{\sin(2\chi_{\max})}{2\pi} \tag{6.26}$$

of the vacuum manifold is covered: the winding number n_W is not quantized. This is a reflection of the fact that whereas topologically nontrivial maps from S^3 to S^3 exist, all maps from R^3 to S^3 can be deformed to the trivial map.

Textures in R^3 are unstable. For the configuration described above, the instability means that $r_c(t) \to 0$ as t increases: the texture collapses. When $r_c(t)$ is microscopical, there will be sufficient energy inside the core to cause $\varphi(0)$ to leave $\mathcal{M}$, pass through 0 and equilibrate at $\chi(0) = \pi$: the texture unwinds.

A further difference compared to topological defects: textures are relevant only for theories with global symmetry. Since all the energy is in spatial gradients, for a local theory the gauge fields can reorient themselves such as to cancel the energy:

$$D_\mu \varphi = 0\,. \tag{6.27}$$

Therefore, it is reasonable to regard textures as an example of a new class of defects, semitopological defects. In contrast to topological defects, there is no core, and $\varphi(\underline{x}) \in \mathcal{M}$ for all $\underline{x}$. In particular, there is no potential energy. In addition, the winding number is not quantized, and hence the defects are unstable. Finally, they exist only in theories with a global internal symmetry.

6.2. Formation

The Kibble mechanism[12)] ensures that in theories which admit topological or semitopological defects, such defects will be produced during a phase transition in the very early Universe.

Consider a mechanical toy model, first introduced by Mazenko, Unruh and Wald[93)] in the context of inflationary Universe models, which is useful in understanding the scalar field evolution. Consider (see Fig. 37) a lattice of points on a

flat table. At each point, a pencil is pivoted. It is free to rotate and oscillate. The tips of nearest neighbor pencils are connected with springs (to mimic the spatial gradient terms in the scalar field Lagrangean). Newtonian gravity creates a potential energy $V(\varphi)$ for each pencil (φ is the angle relative to the vertical direction). $V(\varphi)$ is minimized for $|\varphi| = \eta$ (in our toy model $\eta = \pi/2$). Hence, the Lagrangean of this pencil model is analogous to that of a scalar field with symmetry breaking potential (6.1).

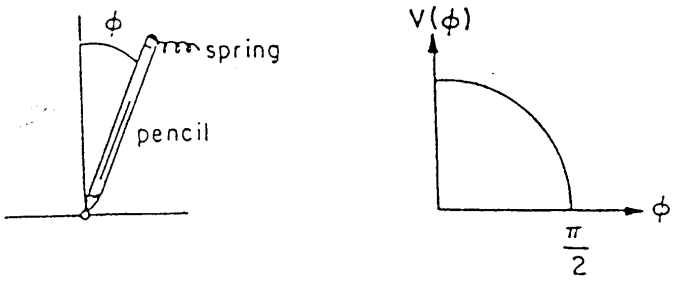

Figure 37: The pencil model: the potential energy of a simple pencil has the same form as that of scalar fields used for spontaneous symmetry breaking. The springs connecting nearest neighbor pencils give rise to contributions to the energy which mimic spatial gradient terms in field theory.

At high temperatures $T \gg T_c$, all pencils undergo large amplitude high frequency oscillations. However, by causality, the phases of oscillation of pencils with large separation s are uncorrelated. For a system in thermal equilibrium, the length s beyond which phases are random is the correlation length $\xi(t)$. However, since the system is quenched rapidly, there is a causality bound on ξ:

$$\xi(t) < t \,, \tag{6.28}$$

where t is the causal horizon.

The critical temperature T_c is the temperature at which the thermal energy is equal to the energy a pencil needs to jump from horizontal to vertical position.

For $T < T_c$, all pencils want to lie flat on the table. However, their orientations are random beyond a distance of $\xi(t)$ determined by equating the free energy gained by symmetry breaking (a volume effect) with the gradient energy lost (a surface effect). As expected, $\xi(T)$ diverges at T_c. Very close to T_c, the thermal energy T is larger than the volume energy gain E_{corr} in a correlation volume. Hence, these domains are unstable to thermal fluctuations. As T decreases, the thermal energy decreases more rapidly than E_{corr}. Below the Ginsburg temperature T_G, there is insufficient thermal energy to excite a correlation volume into the state $\varphi = 0$. Domains of size

$$\xi(t_G) \sim \lambda^{-1}\eta^{-1} \tag{6.29}$$

freeze out[12,142]. The boundaries between these domains become topological defects. An improved version of this argument has recently been given by Zurek[214] (see also Ref. 215).

We conclude that in a theory in which a symmetry breaking phase transitions satisfies the topological criteria for the existence of a fixed type of defect, a network of such defects will form during the phase transition and will freeze out at the Ginsburg temperature. The correlation length is initially given by (6.29), if the field φ is in thermal equilibrium before the transition. Independent of this last assumption, the causality bound implies that $\xi(t_G) < t_G$.

For times $t > t_G$ the evolution of the network of defects may be complicated (as for cosmic strings) or trivial (as for textures). In any case (see the caveats of Refs. 143 and 144), the causality bound persists at late times and states that even at late times, the mean separation and length scale of defects is bounded by $\xi(t) \leq t$.

Applied to cosmic strings, the Kibble mechanism implies that at the time of the phase transition, a network of cosmic strings with typical step length $\xi(t_G)$ will form. According to numerical simulations[145], about 80% of the initial energy is in infinite strings (strings with curvature radius larger than the Hubble radius) and 20% in closed loops.

Note that the Kibble mechanism was discussed above in the context of a global symmetry breaking scenario. As pointed out in Ref. 146, the situation is more complicated in local theories in which gauge field can cancel spatial gradients in φ in the energy functional, and in which spatial gradients in φ can be gauged away. Nevertheless, as demonstrated numerically (in $2+1$ dimensions) in Refs. 42 and 147 and shown analytically in Ref. 148, the Kibble mechanism also applies to local symmetries.

Figure 38: Formation of a loop by a self intersection of an infinite string. According to the original cosmic string scenario, loops form with a radius R determined by the instantaneous coherence length of the infinite string network.

The evolution of the cosmic string network for $t > t_G$ is complicated (see Section 6.4). The key processes are loop production by intersections of infinite strings (see Fig. 38) and loop shrinking by gravitational radiation. These two processes combine to create a mechanism by which the infinite string network loses energy (and length as measured in comoving coordinates). It will be shown (in Section 6.4) that as a consequence, the correlation length of the string network is always proportional to its causality limit

$$\xi(t) \sim t\,. \tag{6.30}$$

Hence, the energy density $\rho_\infty(t)$ in long strings is a fixed fraction of the background energy density $\rho_c(t)$

$$\rho_\infty(t) \sim \mu\xi(t)^{-2} \sim \mu t^{-2} \tag{6.31}$$

or

$$\frac{\rho_\infty(t)}{\rho_c(t)} \sim G\mu\,. \tag{6.32}$$

We conclude that the cosmic string network approaches a "scaling solution" in which the statistical properties of the network are time independent if all distances are scaled to the horizon distance.

Applied to textures, the Kibble mechanism implies that on all scales $r \geq t_G$, field configurations with winding number $n_W \geq n_{cr}$ are frozen in with a probability $p(n_{cr})$ per volume r^3. The critical winding number n_{cr} is defined as the winding number above which field configurations collapse and below which they expand. Only collapsing configurations form clumps of energy which can accrete matter.

The critical winding n_{cr} was determined numerically in Refs. 149 & 150 and analytically in Ref. 151 (see also Ref. 152). It is slightly larger than 0.5. The probability $p(n_{cr})$ can be determined using combinatorial arguments[153].

For $t > t_G$, any configuration on scale $\sim t$ with winding number $n_W \geq n_{cr}$ begins to collapse (before t, the Hubble damping term dominates over the spatial gradient forces, and the field configuration is frozen in comoving coordinates). After unwinding, $\varphi(\underline{x})$ is homogeneous inside the horizon.

The texture model thus also leads to a scaling solution: at all times $t > t_G$ there is the same probability that a texture configuration of scale t will enter the horizon, become dynamical and collapse with a typical time scale t.

6.3. Topological Defects and Cosmology

Topological defects are regions in space with trapped energy density. By Newtonian gravity, these defects can act as seeds about which the matter in the Universe clusters, and hence they play a very important role in cosmology.

As indicated in Table 1, theories with domain walls or with local monopoles are ruled out, and those with only local textures do not give rise to a structure

formation model. As mentioned earlier, theories with domain walls are ruled out since a single wall stretching across the present Universe would overclose it. Local monopoles are also problematic since they do not interact and come to dominate the energy density of the Universe. Local textures do not exist as coherent structures with nonvanishing gradient energy since the gauge fields can always compensate scalar field gradients.

Let us demonstrate explicitly why stable domain walls are a cosmological disaster[140]. If domain walls form during a phase transition in the early Universe, it follows by causality (see however the caveats of Refs. 143 and 144) that even today there will be at least one wall per Hubble volume. Assuming one wall per Hubble volume, the energy density ρ_{DW} of matter in domain walls is

$$\rho_{DW}(t) \sim \eta^3 t^{-1} , \tag{6.33}$$

whereas the critical density ρ_c is

$$\rho_c = H^2 \frac{3}{8\pi G} \sim m_{p\ell}^2 \, t^{-2} . \tag{6.34}$$

Hence, for $\eta \sim 10^{16}$ GeV the ratio of (6.33) and (6.34) is

$$\frac{\rho_{DW}}{\rho_c}(t) \sim \left(\frac{\eta}{m_{p\ell}}\right)^2 (\eta t) \sim 10^{52} . \tag{6.35}$$

The above argument depends in an essential way on the dimension of the defect. One cosmic string per Hubble volume leads to an energy density ρ_{cs} in string

$$\rho_{cs} \sim \eta^2 \, t^{-2} . \tag{6.36}$$

Later in this section we shall see that the scaling (6.36) holds in the cosmic string model. Hence, cosmic strings do not lead to cosmological problems. On the con-

trary, since for GUT models with $\eta \sim 10^{16}$ GeV

$$\frac{\rho_{cs}}{\rho_c} \sim \left(\frac{\eta}{m_{p\ell}}\right)^2 \sim 10^{-6}\,, \tag{6.37}$$

cosmic strings in such theories could provide the seed perturbations responsible for structure formation.

Theories with local monopoles are ruled out on cosmological grounds[141] (see again the caveats of Refs. 143 and 144) for rather different reasons. Since there are no long range forces between local monopoles, their number density in comoving coordinates does not decrease. Since their contribution to the energy density scales as $a^{-3}(t)$, they will come to dominate the mass of the Universe, provided η is sufficiently large.

Theories with global monopoles[154,155] are not ruled out, since there are long range forces between monopoles which lead to a "scaling solution" with a fixed number of monopoles per Hubble volume.

In the following we will describe aspects of two of the promising topological defect models of structure formation, those based on cosmic strings and on global textures. The global monopole scenario is in many aspects similar to the texture theory.

6.4. Cosmic String Evolution and Scaling

If the evolution of the cosmic string network were trivial in the sense that all strings would only stretch as the universe expands, there would be an immediate cosmological disaster. Consider a fixed comoving volume V with a string passing through. The energy in radiation decreases as $a^{-1}(t)$ while the energy in string increases as $a(t)$. Hence trivial evolution would immediately lead to a string dominated universe, a cosmological disaster. In order to study the evolution of a cosmic string network, it is neccessary to know the effective action for a string, and to study what happens when two strings cross.

The equations of motion of a string are determined by the Nambu action

$$S = -\mu \int d\sigma d\tau \left(- \det g^{(2)}_{ab} \right)^{1/2} \quad a, b = 0, 1 \tag{6.38}$$

where $g^{(2)}_{ab}$ is the world sheet metric and σ and τ are the world sheet coordinates. In flat space-time, τ can be taken to be coordinate time, and σ is an affine parameter along the string. In terms of the string coordinates $X^\mu(\sigma, \tau)$ and the metric $g^{(4)}_{\mu\nu}$ of the background space-time,

$$g^{(2)}_{ab} = X^\mu_{,a} X^v_{,b} g^{(4)}_{\mu\nu} \, . \tag{6.39}$$

From general symmetry considerations, it is possible to argue that the Nambu action is the correct action. However, I shall follow Foerster[156] and Turok[157] and give a direct heuristic derivation. We start from a general quantum field theory Lagrangean $\mathcal{L}_{QFT}$. The action is

$$S = \int d^4 y \mathcal{L}_{QFT} \left(\varphi(y) \right) \tag{6.40}$$

We assume the existence of a linear topological defect at $X^\mu(\sigma, \tau)$. The idea now is to change variables such that σ and τ are two of the new coordinates, and to expand S to lowest order in w/R, where w is the width of the string and R its curvature radius. As the other new coordinates we take the coordinates ρ^2 and ρ^3 in the normal plane to $X^\mu(\sigma, \tau)$. Thus the transformation takes the old coordinates $y^\mu (\mu = 0, 1, 2, 3)$ to new ones $\sigma^a = (\tau, \sigma, \rho^2, \rho^3)$:

$$y^\mu(\sigma^a) = X^\mu(\sigma, \tau) + \rho^i n^\mu_i(\sigma, \tau) \tag{6.41}$$

where $i = 2, 3$ and n^μ_i are the basis vectors in the normal plane to the string world

sheet. The measure transforms as

$$\int d^4y = \int d\sigma d\tau d\rho^2 d\rho^3 (\det M^\mu_a) \tag{6.42}$$

with

$$M^\mu_a = \frac{\partial y^\mu}{\partial \sigma^a} = \begin{pmatrix} \partial X^\mu/\partial(\sigma,\tau) \\ n^\mu_i \end{pmatrix} + O(\rho)\,. \tag{6.43}$$

The determinant can easily be evaluated using the following trick

$$\det M^\mu_a = (-\det \eta_{\mu\nu} M^\mu_a M^\nu_b)^{1/2} \equiv \sqrt{-\det D_{ab}} \tag{6.44}$$

$$D = \begin{pmatrix} \frac{\partial x^\mu}{\partial(\sigma,\tau)} \frac{\partial X^\nu}{\partial(\sigma,\tau)} \eta_{\mu\nu} & \frac{\partial X^\mu}{\partial(\sigma,\tau)} n^\nu_b \eta_{\mu\nu} \\ \frac{\partial X^\mu}{\partial(\sigma,\tau)} n^\nu_a \eta_{\mu\nu} & n^\mu_a n^\nu_b \eta_{\mu\nu} \end{pmatrix} = \begin{pmatrix} X^\mu_{,a} X^\nu_{,b} \eta_{\mu\nu} & 0 \\ 0 & \delta_{ab} \end{pmatrix} + 0\left(\frac{w}{R}\right) \tag{6.45}$$

Hence

$$\begin{aligned} S &= \int d\sigma d\tau \left(-\det g^{(2)}_{ab}\right)^{1/2} \int d\rho^2 d\rho^3 \mathcal{L}(y(\sigma,\tau,\rho^2,\rho^3)) + O\left(\frac{w}{R}\right) \\ &= -\mu \int d\sigma d\tau \left(-\det g^{(2)}_{ab}\right)^{1/2} + O\left(\frac{w}{R}\right)\,. \end{aligned} \tag{6.46}$$

$-\mu$ is the integral of $\mathcal{L}$ in the normal plane of X. To first order in w/R, it equals the integral of $-\mathcal{H}$; hence it is the mass per unit length.

This derivation of the Nambu action is instructive as it indicates a method for calculating corrections to the equations of motion of the string when extra fields are present, *e.g.* for superconducting cosmic strings. It also gives a way of calculating the finite thickness corrections to the equations of motion which will be important at cusps (see below).

In flat space-time we can consistently choose $\tau = t, \dot{x}\cdot x' = 0$ and $\dot{x}^2 + x'^2 = 0$. The equations of motion derived from the Nambu action then become

$$\ddot{\underline{\mathrm{x}}} - \underline{\mathrm{x}}'' = 0\,. \tag{6.47}$$

where $'$ indicates the derivative with respect to σ. The general solution can be

decomposed into a left moving and a right moving mode[158)]

$$\underline{\mathrm{x}}(t,\sigma) = \frac{1}{2}\left[\underline{\mathrm{a}}(\sigma - t) + \underline{\mathrm{b}}(\sigma + t)\right] \tag{6.48}$$

The gauge conditions imply

$$\dot{\underline{\mathrm{a}}}^2 = \dot{\underline{\mathrm{b}}}^2 = 1 \tag{6.49}$$

For a loop, $\underline{\mathrm{x}}(\sigma, t)$ is periodic and hence the time average of $\dot{\underline{\mathrm{a}}}$ and $\dot{\underline{\mathrm{b}}}$ vanish. $\dot{\underline{\mathrm{a}}}$ and $\dot{\underline{\mathrm{b}}}$ are hence closed curves on the unit sphere with vanishing average. Two such curves generically intersect if they are continuous. An intersection corresponds to a point with $\underline{\mathrm{x}}' = 0$ and $\dot{\underline{\mathrm{x}}} = 1$. Such a point moving at the speed of light is called a cusp. $\dot{\underline{\mathrm{x}}}(\sigma, t)$ need not be continuous. Points of discontinuity are called kinks. Note that both cusps and kinks will be smoothed out by finite thickness effects[159)].

The Nambu action does not describe what happens when two strings hit. This process has been studied numerically for both global[160)] and local[161)] strings. The authors of these papers set up scalar field configurations corresponding to two strings approaching one another and evolve the complete classical scalar field equations. The result of the analysis is that strings do not cross but exchange ends, provided the relative velocity is smaller than 0.9. Thus, by self intersecting, an infinite string will split off a loop (Fig. 38). An important open problem is to understand this process analytically. For a special value of the coupling constant Ruback[162)] has given a mathematical explanation (see also Shellard and Ruback in Ref. 161).

There are two parts to the nontrivial evolution of the cosmic string network. Firstly, loops are produced by self intersections of infinite strings. Loops oscillate due to the tension and slowly decay by emitting gravitational radiation. Combining the two steps we have a process by which energy is transferred from the cosmic string network to radiation.

There are analytical indications that a stable "scaling solution" (already described in Section) for the cosmic string network exists. In the scaling solution, on

the order of 1 infinite string segment crosses every Hubble volume. The correlation length $\xi(t)$ of an infinite string is thus of the order t. A heuristic argument for the scaling solution is due to Vilenkin[5]. Take $\tilde{\nu}(t)$ to be the mean number of infinite string segments per Hubble volume. Then the energy density in infinite strings is

$$\rho_\infty(t) = \mu\tilde{\nu}(t)t^{-2} \tag{6.50}$$

The number of loops $n(t)$ produced per unit volume is proportional to the square of $\tilde{\nu}$, since it takes two string segments to generate a string intersection. Hence,

$$\frac{dn(t)}{dt} = c\tilde{\nu}^2 t^{-4} \tag{6.51}$$

where c is a constant of the order 1. Conservation of energy in strings gives

$$\frac{d\rho_\infty(t)}{dt} + \frac{3}{2t}\,\rho_\infty(t) = -c'\mu t\,\frac{dn}{dt} = -c'\mu\tilde{\nu}^2 t^{-3} \tag{6.52}$$

or, written as an equation for $\tilde{\nu}(t)$

$$\dot{\tilde{\nu}} - \frac{\tilde{\nu}}{2t} = -cc'\tilde{\nu}^2 t^{-1} \tag{6.53}$$

Thus if $\tilde{\nu} \gg 1$ then $\dot{\tilde{\nu}} < 0$ while if $\tilde{\nu} \ll 1$ then $\dot{\tilde{\nu}} > 0$. Hence there will be a stable solution with $\tilde{\nu} \sim 1$.

The precise value of $\tilde{\nu}$ must be determined in numerical simulations. These simulations are rather difficult because of the large dynamic range required and due to singularities which arise in the evolution equations near cusps. In the radiation dominated epoch, $\tilde{\nu}$ is still uncertain by a factor of about 10. The first results were reported in Ref. 163. More recent results are due three groups. Bennett and Bouchet[164] and Allen and Shellard[165] are converging on a value $10 < \tilde{\nu} < 20$, whereas Albrecht and Turok[166] obtain a value which is about 100.

The scaling solution for the infinite strings implies that the network of strings looks the same at all times when scaled to the Hubble radius. This should also imply that the distribution of cosmic string loops is scale invariant in the same sense. At present, however, there is no convincing evidence from numerical simulations that this is really the case.

A scaling solution for loops implies that the distribution of $R_i(t)$, the radius of loops at the time of formation, is time independent after dividing by t. To simplify the discussion, I shall assume that the distribution is monochromatic, *i.e.*

$$R_i(t)/t = \alpha\,. \tag{6.54}$$

Based on Fig. 38, we expect $\alpha \sim 1$. The numerical simulations[164–166], however, now give $\alpha < 10^{-2}$. This is due to the fact that there is a lot of small scale structure on the long strings, and that the typical scale of loop production is not determined by the overall curvature radius of the long strings, but rather by the typical lengths of the small scale structure.

From the scaling solution (6.50) for the infinite strings we can derive the scaling solution for loops. We assume that the energy density in long strings – inasmuch as it is not redshifted – must go into loops. β shall be a measure for the mean length ℓ in a loop of "radius" R

$$\ell = \beta R\,. \tag{6.55}$$

If per expansion time and Hubble volume about 1 loop of radius $R_i(t)$ is produced, then we know that the number density in physical coordinates of loops of radius $R_i(t)$ is

$$n(R_i(t), t) = ct^{-4} \tag{6.56}$$

with a constant c which can be calculated from (6.50), (6.54) and (6.55). Neglecting

gravitational radiation, this number density simply redshifts

$$n(R,t) = \left(\frac{z(t)}{z(t_f(R))}\right)^3 n(R, t_f(R))\,, \tag{6.57}$$

where $t_f(R)$ is the time when loops of radius R are formed. Isolating the R dependence, we obtain

$$n(R,t) \sim R^{-4} z(R)^{-3} \tag{6.58}$$

where $z(R)$ is the redshift at time $t = R$. We have the following special cases:

$$\begin{aligned} n(R,t) &\sim R^{-5/2} t^{-3/2} \qquad && t < t_{eq} \\ n(R,t) &\sim R^{-5/2} t_{eq}^{1/2} t^{-2} \qquad && t > t_{eq}\,,\ t_f(R) < t_{eq} \\ n(R,t) &\sim R^{-2} t^{-2} \qquad && t > t_{eq}\,,\ t_f(R) > t_{eq}\,. \end{aligned} \tag{6.59}$$

The proportionality constant c is

$$c = \frac{1}{2}\beta^{-1}\alpha^{-2}\tilde{\nu} \tag{6.60}$$

(see *e.g.* Ref. 167). In deriving (6.60) it is important to note that $n(R_i(t), t)dR_i$ is the number density of loops in the radius interval $[R_i, R_i + dR_i]$. Hence, in the radiation dominated epoch

$$n(R,t) = \nu R^{-5/2} t^{-3/2} \tag{6.61}$$

with

$$\nu = \frac{1}{2}\beta^{-1}\alpha^{1/2}\tilde{\nu}\,. \tag{6.62}$$

From (6.62) we can read off the uncertainties in ν based on the uncertainties in the numerical results. Both $\alpha^{1/2}$ and $\tilde{\nu}$ are determined only up to one order of magnitude. Hence, any quantitative results which depend on the exact value of ν are rather uncertain.

Gravitational radiation leads to a lower cutoff in $n(R,t)$. Loops with radius smaller than this cutoff were all formed at essentially the same time and hence have the same number density. Thus, $n(R)$ becomes flat. The power in gravitational radiation P_G can be estimated using the quadrupole formula[168]. For a loop of radius R and mass M

$$P_G = \frac{1}{5} G < \dddot{Q}\dddot{Q} > , \tag{6.63}$$

where Q is the quadrupole moment, $Q \sim MR^2$, and since the frequency of oscillation is $\omega = R^{-1}$

$$P_G \sim G(MR^2)^2\omega^6 \sim (G\mu)\mu \, . \tag{6.64}$$

Even though the quadrupole approximation breaks down since the loops move relativistically, (6.64) gives a good order of magnitude of the power of gravitational radiation. Improved calculations give[169]

$$P_G = \gamma(G\mu)\mu \tag{6.65}$$

with $\gamma \sim 50$. (6.55) and (6.65) imply that

$$\dot{R} = \tilde{\gamma} G\mu \tag{6.66}$$

with $\tilde{\gamma} \equiv \gamma/\beta \sim 5$ (using $\beta \simeq 10$). Note that the rate of decrease is constant. Hence,

$$R(t) = R_i - (t - t_i)\tilde{\gamma} G\mu \tag{6.67}$$

and the cutoff loop radius is

$$R_c \sim \tilde{\gamma} G\mu t_i \, . \tag{6.68}$$

Let us briefly summarize the scaling solution

1) At all times the network of infinite strings looks the same when scaled by the Hubble radius. A small number of infinite string segments cross each Hubble volume and $\rho_\infty(t)$ is given by (6.50).

2) There is a distribution of loops of all sizes $0 \leq R < t$. Assuming scaling for loops, then

$$n(R,t) = \nu R^{-4} \left(\frac{z(t)}{z(R)}\right)^3 \,,\; R\,\epsilon\,[\tilde{\gamma}G\mu t, \alpha t] \tag{6.69}$$

where $\alpha^{-1}R$ is the time of formation of a loop of radius R. Also

$$n(R,t) = n(\tilde{\gamma}G\mu t, t)\,,\; R < \tilde{\gamma}G\mu t\,. \tag{6.70}$$

Although the qualitative characteristics of the cosmic string scaling solution are well established, the quantitative details are not. The main reason for this is the fact that the Nambu action breaks down at kinks and cusps. However, kinks and cusps inevitably form and are responsible for the small scale structure on strings. In fact, coarse graining by integrating out the small scale structure may give an equation of state for strings which deviates from that of a Nambu string[170]. Attempts at understanding the small scale structure on strings are at present under way[171].

6.5. Cosmic Strings and Structure Formation

The starting point of the structure formation scenario in the cosmic string theory is the scaling solution for the cosmic string network, according to which at all times t (in particular at t_{eq}, the time when perturbations can start to grow) there will be a few long strings crossing each Hubble volume, plus a distribution of loops of radius $R \ll t$ (see Fig. 39).

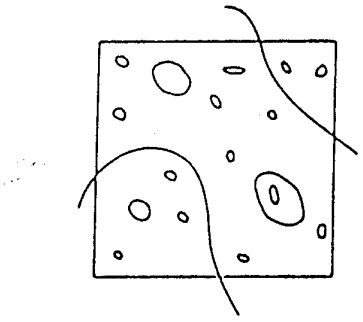

Figure 39. Sketch of the scaling solution for the cosmic string network. The box corresponds to one Hubble volume at arbitrary time t.

The cosmic string model admits three mechanisms for structure formation: loops, filaments, and wakes. Cosmic string loops have the same time averaged field as a point source with mass[172)]

$$M(R) = \beta R \mu \,, \tag{6.71}$$

R being the loop radius and $\beta \sim 2\pi$. Hence, loops will be seeds for spherical accretion of dust and radiation.

For loops with $R \leq t_{eq}$, growth of perturbations in a model dominated by cold dark matter starts at t_{eq}. Hence, the mass at the present time will be

$$M(R,\, t_0) = z(t_{eq}) \beta \, R \mu \,. \tag{6.72}$$

In the original cosmic string model[47,48,57)] it was assumed that loops dominate over wakes. In this case, the theory could be normalized (*i.e.*, μ could be determined) by demanding that loops with the mean separation of clusters d_{cl} (from the discussion in Section 6.4 it follows that the loop radius $R(d_{cl})$ is determined by the mean separation) accrete the correct mass, *i.e.*, that

$$M(R(d_{cl}), t_0) = 10^{14} M_\odot \,. \tag{6.73}$$

This condition yields[57)]

$$\mu \simeq 10^{32}\mathrm{GeV}^2 \tag{6.74}$$

Thus, if cosmic strings are to be relevant for structure formation, they must arise due to a symmetry breaking at energy scale $\eta \simeq 10^{16}$GeV. This scale happens to be the scale of unification (GUT) of weak, strong and electromagnetic interactions. It is tantalizing to speculate that cosmology is telling us that there indeed was new physics at the GUT scale.

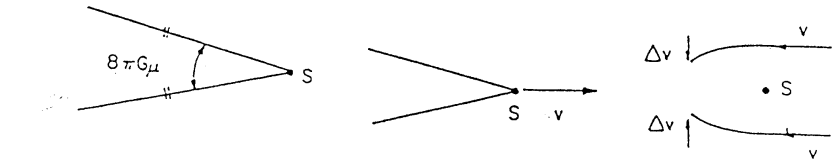

Figure 40. Sketch of the mechanism by which a long straight cosmic string S moving with velocity v in transverse direction through a plasma induces a velocity perturbation Δv towards the wake. Shown on the left is the deficit angle, in the center is a sketch of the string moving in the plasma, and on the right is the sketch of how the plasma moves in the frame in which the string is at rest.

The second mechanism involves long strings moving with relativistic speed in their normal plane which give rise to velocity perturbations in their wake[173)]. The mechanism is illustrated in Fig. 40: space normal to the string is a cone with deficit angle[174)]

$$\alpha = 8\pi G\mu\,. \tag{6.75}$$

If the string is moving with normal velocity v through a bath of dark matter, a velocity perturbation

$$\delta v = 4\pi G\mu v\gamma \tag{6.76}$$

[with $\gamma = (1-v^2)^{-1/2}$] towards the plane behind the string results. At times after t_{eq}, this induces planar overdensities, the most prominent (*i.e.*, thickest at the

present time) and numerous of which were created at t_{eq}, the time of equal matter and radiation[58,59,63]. The corresponding planar dimensions are (in comoving coordinates)

$$t_{eq}z(t_{eq}) \times t_{eq}z(t_{eq})v \sim (40 \times 40v)\,\mathrm{Mpc}^2\,. \tag{6.77}$$

The thickness d of these wakes can be calculated using the Zel'dovich approximation[63]. The result is

$$d \simeq G\mu v\gamma(v)z(t_{eq})^2\,t_{eq} \simeq 4v\,\mathrm{Mpc}\,. \tag{6.78}$$

Wakes arise if there is little small scale structure on the string. In this case, the string tension equals the mass density, the string moves at relativistic speeds, and there is no local gravitational attraction towards the string.

In contrast, if there is small scale structure on strings, then the string tension T is smaller[170] than the mass per unit length μ and the metric of a string in z direction becomes[175]

$$ds^2 = (1 + h_{00})(dt^2 - dz^2 - dr^2 - (1 - 8G\mu)r^2dy^2) \tag{6.79}$$

with

$$h_{00} = 4G(\mu - T)\ln\frac{r}{r_0}\,, \tag{6.80}$$

r_0 being the string width. Since h_{00} does not vanish, there is a gravitational force towards the string which gives rise to cylindrical accretion, thus producing filaments.

As is evident from the last term in the metric (6.79), space perpendicular to the string remains conical, with deficit angle given by (6.75). However, since the string is no longer relativistic, the transverse velocities v of the string network are expected to be smaller, and hence the induced wakes will be shorter and thinner.

Which of the mechanisms – filaments or wakes – dominates is determined by the competition between the velocity induced by h_{00} and the velocity perturbation of the wake. The total velocity is[175)]

$$u = -\frac{2\pi G(\mu - T)}{v\gamma(v)} - 4\pi G\mu v\gamma(v)\,, \tag{6.81}$$

the first term giving filaments, the second producing wakes. Hence, for small v the former will dominate, for large v the latter.

By the same argument as for wakes, the most numerous and prominent filaments will have the distinguished scale

$$t_{eq} z(t_{eq}) \times d_f \times d_f \tag{6.82}$$

where d_f can be calculated using the Zel'dovich approximation[216)].

The cosmic string model predicts a scale-invariant spectrum of density perturbations, exactly like inflationary Universe models but for a rather different reason. Consider the *r.m.s.* mass fluctuations on a length scale $2\pi k^{-1}$ at the time $t_H(k)$ when this scale enters the Hubble radius. From the cosmic string scaling solution it follows that a fixed (*i.e.*, $t_H(k)$ independent) number $\tilde{v}$ of strings of length of the order $t_H(k)$ contribute to the mass excess $\delta M(k,\, t_H(k))$. Thus

$$\frac{\delta M}{M}\,(k,\, t_H(k)) \sim \frac{\tilde{v}\mu t_H(k)}{G^{-1}t_H^{-2}(k)t_H^3(k)} \sim \tilde{v}\, G\mu\,. \tag{6.83}$$

Note that the above argument predicting a scale invariant spectrum will hold for all topological defect models which have a scaling solution, in particular also for global monopoles and textures.

The amplitude of the *r.m.s.* mass fluctuations (equivalently: of the power

spectrum) can be used to normalize $G\mu$. Since today on galaxy cluster scales

$$\frac{\delta M}{M}(k,\, t_0) \sim 1\,, \tag{6.84}$$

the growth rate of fluctuations linear in $a(t)$ yields

$$\frac{\delta M}{M}\,(k,\, t_{eq}) \sim 10^{-4}\,, \tag{6.85}$$

and therefore, using $\tilde{v} \sim 10$,

$$G\mu \sim 10^{-5}\,. \tag{6.86}$$

A big advantage of the cosmic string model over inflationary Universe models is that HDM is a viable dark matter candidate. Cosmic string loops survive free streaming, as discussed in Section 3.4, and can generate nonlinear structures on galactic scales, as discussed in detail in Refs. 61 and 62. Accretion of hot dark matter by a string wake was studied in Ref. 63. In this case, nonlinear perturbations develop only late. At some time t_{nl}, all scales up to a distance $q_{\max}$ from the wake center go nonlinear. Here

$$q_{\max} \sim G\mu v\gamma(v) z(t_{eq})^2 t_{eq} \sim 4v\,\mathrm{Mpc}\,, \tag{6.87}$$

and it is the comoving thickness of the wake at t_{nl}. Demanding that t_{nl} corresponds to a redshift greater than 1 leads to the constraint

$$G\mu > 5 \cdot 10^{-7}\,. \tag{6.88}$$

Note that in a cosmic string and hot dark matter model, wakes form nonlinear structures only very recently. Accretion onto loops and small scale structure on the long strings provide two mechanisms which may lead to high redshift objects such as quasars and high redshift galaxies. The first mechanism has recently been studied in Ref. 217.

The power spectra in the cosmic string models with CDM and HDM are obviously different on scales smaller than the maximal neutrino free streaming length (3.27). Recent calculations[176,177] of the power spectra are shown in Fig. 41.

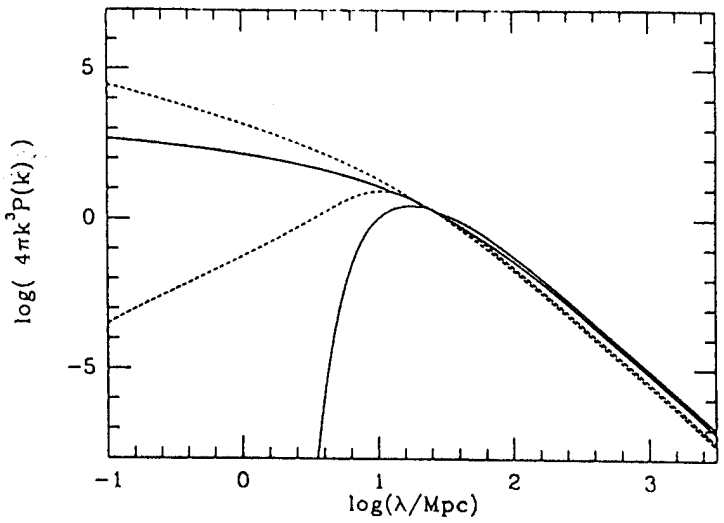

Figure 41: Power spectra for cosmic string HDM and CDM theories (dashed curves), compared to those for inflationary HDM and CDM models (solid curves). In each case, the top curve is for CDM, the bottom one for HDM. Note that there is substantial power on small scales in the cosmic string HDM theory.

6.6. Global Textures and Structure Formation

The starting point of the texture scenario of structure formation[60] is the scaling solution for textures: at any time t, there is a fixed probability $p(n_w)\,dn_w$ that the scalar field configuration over a Hubble volume covers between n_w and $n_w + dn_w$ of the vacuum manifold, *i.e.*, we have a texture with winding number in the interval $[n_w,\, n_w + dn_w]$ entering the Hubble radius.

The dynamics of a texture is easy to understand. Consider the spherically symmetric texture configuration of (6.24) with $\chi(r)$ increasing from 0 to $\chi_{\max}$ over a distance d. If d is larger than the Hubble radius, then the Hubble damping

term dominates the equation of motion for φ and the field configuration is frozen in. Once the Hubble radius t catches up with d, the microphysical forces become dominant and the texture field begins to evolve.

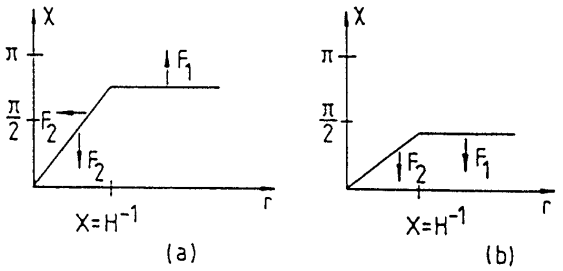

Figure 42: A sketch of the forces acting on a spherically symmetric texture configuration and which cause unwinding in case (a) in which the winding number is larger than the critical winding, and dissipation if the winding is smaller than its critical value (case (b)). r is the distance from the center of the texture, and the vertical axis shows the value of the χ field.

The evolution of φ tends to minimize the field energy. Consider first large distances from the texture center. The spatial gradient energy can be decreased by having $\chi_{\max}$ increase (if $\chi_{\max} > \chi_c$) or decrease (if $\chi_{\max} < \chi_c$) (see Figure 42). The winding associated with χ_c is called the critical winding n_c (see (6.26)). For a single texture in an infinite volume we would expect

$$\chi_c = \frac{\pi}{2} \, (i.e., \, n_c = 0.5) \, . \tag{6.89}$$

For realistic textures there will be a "finite volume cutoff" determined by the separation of textures. A semi-analytical analysis and numerical simulations give[149–151)]

$$n_c \simeq 0.6 \, . \tag{6.90}$$

If $n_W < n_c$, then the field configuration will relax to a trivial one. No localized energy concentrations will be generated, and we cannot speak of a "texture."

However, if $n_W > n_c$ the field evolution will be more interesting. At large r, $\chi(r)$ will increase. In addition, the radius $r(\chi)$ where χ takes on a fixed value χ tends to decrease, since this leads to a concentration of gradient energies over a smaller region. Hence, the field configuration will contract (see Fig. 42), with increasing total winding number. Eventually, close to $r = 0$ there is sufficient tension energy for φ to be able to leave the vacuum manifold and jump from $\chi = 0$ to $\chi = \pi$. This is the texture unwinding event. After unwinding, energy is radiated radially in the form of Goldstone bosons.

In the texture model it is the contraction of the field configuration which leads to density perturbations[178]. At the time when the texture enters the horizon, an isocurvature perturbation is established: the energy density in the scalar field is compensated by a deficit in radiation. However, the contraction of the scalar field configuration leads to a clumping of gradient and kinetic energy at the center of the texture (Fig. 43). This, in turn, provides the seed perturbations which cause dark matter and radiation to collapse in a spherical manner[179,180].

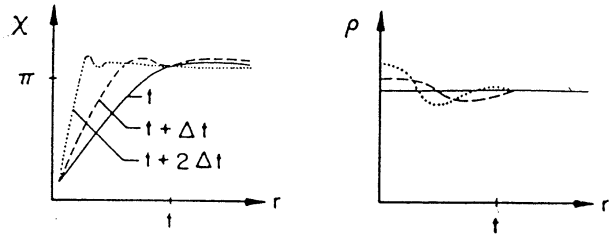

Figure 43: A sketch of the density perturbation produced by a collapsing texture. The left graph shows the time evolution of the field $\chi(r)$ as a function of radius r and time (see (5.18)). The contraction of $\chi(r)$ leads to a spatial gradient energy perturbation at the center of the texture, as illustrated on the right. The energy is denoted by ρ. Solid lines denote the initial time, dashed lines are at time $t+\Delta t$, and dotted lines correspond to time $t + 2\Delta t$, where Δt is a fraction of the Hubble expansion time (the typical time scale for the dynamics).

As in the cosmic string model, also in the global texture scenario the length

scale of the dominant structures is the comoving Hubble radius at t_{eq}. Textures generated at t_{eq} are the most numerous, and the perturbations induced by them have the most time to grow.

As mentioned in the previous subsection, the texture model predicts a scale-invariant spectrum of density perturbations. Hence, in order to differentiate topological defect models from inflationary scenarios, and to distinguish between different topological defect theories, we need statistics which are not determined by the power spectrum alone. We need statistics which are sensitive to the non-random phases of topological defect models. One such statistic is the genus curve[181]. For a surface S embedded in R^3, the genus g is

$$g(S) = \#\text{ of holes of } S - \#\text{ of disconnected components of } S + 1\,. \tag{6.91}$$

The genus g can now be evaluated for the isodensity surface $S(\rho)$, the surface of points in space with density equal to ρ. The curve

$$g(\rho) = g(S(\rho)) \tag{6.92}$$

is the genus curve. To reduce numerical errors, g can also be evaluated based on a cell decomposition of the volume. Now, $g(n)$ is the genus of the boundary of the cell complex in which each cell contains more than n galaxies. In this case, the genus is simply

$$g = 1 - \frac{1}{2}(V - E - F) \tag{6.93}$$

where V, E, F are the number of vertices, edges, and faces of the polygonal surface, respectively.

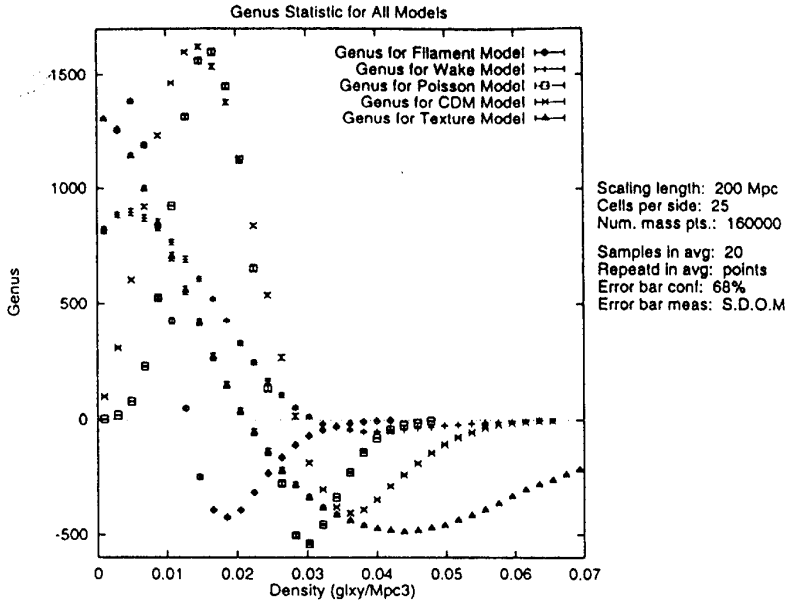

Figure 44: Comparison of the genus curve (genus as a function of galaxy density) of different toy models of structure formation. Except for the Gaussian model, all theories have the same linear power spectrum. The 'filament', 'wake' and 'texture' toy models are based on laying down at random linear, planar and spherical overdense regions of galaxies. Thus, the figure demonstrates that the genus statistic is able to distinguish between theories with different topologies but identical power spectra. The 'CDM' model predictions are computed from linear theory, and the 'Poisson' model is obtained by randomly distributing galaxies. See the senior thesis by Aguirre[182)] for further details.

As shown in Fig. 44, the genus statistic is able to distinguish between models with the same power spectrum but different phase correlations and topology[182)]. For a texture toy model, the genus curve is mostly negative, for a cosmic string wake model, it is predominately positive. The differences compared to a random phase inflationary model are statistically significant.

The differences shown in Fig. 44 will only be apparent in large-scale samples of galaxies, *i.e.*, on scales exceeding the comoving radius at t_{eq}. Such samples should,

however, become available in the near future, and at that point genus curve and other statistics sensitive to non-random phases should become a powerful tool for distinguishing the predictions of the different models of structure formation.

A final word concerning textures: since they are short-lived, only CDM is a viable dark matter candidate in the context of this structure formation scenario.

7. Cosmic Microwave Background Anisotropies

As mentioned in Section 3, the near-isotropy of the CMB is the strongest evidence in support of the cosmological principle. By the same reasoning, any density inhomogeneities in the early Universe will give rise to CMB anisotropies. Since our present theories of galaxy formation are based on the gravitational instability scenario, they predict such inhomogeneities. The CMB temperature fluctuations probe the structure of space at t_{rec}, the time of last scattering, a time when the density perturbations still have a small amplitude and can be analyzed in linear theory. Hence, a study of CMB anisotropies will yield a lot of constraints for structure formation models. The information gained will be robust, *i.e.*, independent of the uncertainties of nonlinear gravitational and hydrodynamical effects, but it will deal only with large scales (comparable or larger than the comoving horizon at t_{rec}). In this section we shall give a brief overview of the theory of CMB anisotropies and summarize some recent observational results.

7.1. Basics

As illustrated in Fig. 45, there are three main sources of CMB anisotropies. The first are gravitational potential perturbations at t_{rec} which lead to fluctuations of the surface of last scattering. This produces deviations in the light travel time between last scattering and detection, and – given that the photons have the same temperature on the surface of last scattering – to temperature fluctuations for the observer.

Figure 45: Space-time plot sketching the origin of CMB temperature anisotropies. The surface labelled T_{rec} is the last scattering surface. $\mathcal{O}$ is the observer at the present time measuring photons γ impinging from directions in the sky separated by an angle θ. The shaded area labelled C is the world volume of a local overdensity, leading to distortions of geodesics. Possible velocities of observer and emitter are indicated by $\vec{v}_o$ and $\vec{v}_e$, respectively.

The second source is due to gravitational perturbations along the line of sight which lead to deviations of the geodesics and hence to temperature differences. A Newtonian way of understanding this effect is to consider a photon passing through a large mass concentration. On the way towards the center, the photon is falling into a potential well and acquires a blueshift, whereas on its way out it is redshifted. In an expanding background, this redshift does not exactly cancel the initial blueshift, and a temperature fluctuation results.

The third source contributing to CMB anisotropies are peculiar velocities on the surface of last scattering and of the observer. The peculiar motion of the earth gives rise to a dipole anisotropy[183)]

$$\frac{\delta T}{T}\Big|_{\text{dipole}} \simeq 10^{-3} \tag{7.1}$$

Peculiar velocities induce temperature fluctuations by means of the Doppler effect.

For linear adiabatic density perturbations in a matter dominated Universe, the line of sight contributions to $\delta T/T$ can be written as total time derivative and thus reduces to a contribution from the surface of last scattering and can be simply combined with the potential fluctuations at t_{rec}. This is the case first studied by Sachs and Wolfe, and the combined effect is now called the Sachs-Wolfe effect[184)].

An analysis of the Sachs-Wolfe effect reveals a very simple relationship between temperature fluctuations $\delta T/T\,(\vartheta)$ on an angular scale ϑ and the magnitude of density perturbations on the corresponding lengths scale $\lambda(\vartheta)$, where at last

scattering $\lambda(\vartheta)$ equals the distance subtended by two light rays with angular separation ϑ (see Fig. 45). A simple derivation[11,185] of this relationship makes use of the gauge invariant theory of cosmological perturbations described in Section 4.4.

The starting point is the phase space distribution function $f(x^\alpha, p_i)$ which would be a function of p/T exclusively in the absence of inhomogeneities. In the presence of inhomogeneities, the deviation of f from homogeneity is associated with temperature fluctuations:

$$f(x^\alpha, p_i) = \bar{f}(p/\bar{T} + \delta T)\,, \tag{7.2}$$

where $\bar{T}$ is the average temperature and $\bar{f}(p/T)$ is the background phase space density.

The phase space distribution function satisfies the collisionless Boltzmann equation

$$\frac{dx^\alpha}{d\eta}\,\frac{\partial f}{\partial x^\alpha} + \frac{dp_i}{d\eta}\,\frac{\partial f}{\partial p_i} = 0 \tag{7.3}$$

where, as in Section 4.4, the variable η denotes conformal time. This equation can be integrated along the perturbed geodesics which are given by

$$\frac{dp_\alpha}{d\eta} = 2p\frac{\partial \Phi}{\partial x^\alpha} \tag{7.4}$$

and

$$\frac{dx^i}{d\eta} = l^i(1 + 2\Phi), \tag{7.5}$$

with

$$l^i = -\frac{1}{p}\,p_i \tag{7.6}$$

and

$$p^2 = p_i p_i\,. \tag{7.7}$$

Inserting these relations into the Boltzman equation gives

$$\left(\frac{\partial}{\partial\eta} + l^i\partial_i\right)\frac{\delta T}{T} = -2l^i\partial_i\Phi\,. \tag{7.8}$$

Since in the matter dominated period $\partial_\eta\Phi = 0$ we can rewrite (7.8) as

$$\left(\frac{\partial}{\partial_\eta} + l^i_i\right)\left(\frac{\delta T}{T} + 2\Phi\right) = 0\,, \tag{7.9}$$

which implies that

$$\frac{\delta T}{T} + 2\Phi = \text{ const} \tag{7.10}$$

along the perturbed geodesics.

For isothermal primordial perturbations ($\frac{\delta T}{T}\,(t_{rec}) = 0$), the result (7.10) implies that

$$\frac{\delta T}{T}(\eta_0) = 2\Phi(\eta_{rec}) + l^i v_i(\eta_{rec}) \tag{7.11}$$

whereas for primordial adiabatic perturbations (vanishing initial entropy perturbations)

$$\frac{\delta T}{T}(\eta_0) = \frac{1}{3}\Phi(\eta_{rec}) + l^i v_i(\eta_{rec})\,, \tag{7.12}$$

i.e., the combination of initial curvature fluctuations and line of sight effects leads to a partial cancellation of the anisotropy. The second term on the r.h.s. of (7.11) and (7.12) is the Doppler term, and it arises from a determination of the constant in (7.10) based on considering the initial conditions at t_{rec}.

Since Φ is constant both between t_{eq} and t_{rec} and while outside the Hubble radius, and since

$$\Phi(t_H) \sim \frac{\delta\rho}{\rho}\,(t_H) \tag{7.13}$$

at Hubble radius crossing t_H, our results imply that (modulo Doppler terms) for

adiabatic perturbations

$$\frac{\delta T}{T}(\vartheta,\, t_0) = \frac{1}{3}\Phi(\lambda(\vartheta),\, t_{eq}) \sim \frac{1}{3}\,\frac{\delta M}{M}\,(\lambda(\vartheta),\, t_H))\,. \tag{7.14}$$

We conclude that the spectrum of primordial mass perturbations can be normalized by CMB anisotropy detections. For a scale invariant spectrum of density perturbations, the *r.m.s.* temperature fluctuations are predicted to be independent of ϑ on angular scales larger than the Hubble radius at t_{rec} (between 1 and 2 degrees).

7.2. Specific Signatures

All theories of structure formation give rise to Sachs-Wolfe type temperature fluctuations given by (7.13) and (7.14). In topological defect models there are, in addition, specific signatures which cannot be described in a linear perturbative analysis.

As described in Section 6.5, space perpendicular to a long straight cosmic string is conical with deficit angle given by (6.75). Consider now CMB radiation approaching an observer in a direction normal to the plane spanned by the string and its velocity vector (see Fig. 46). Photons arriving at the observer having passed on different sides of the string will obtain a relative Doppler shift which translates into a temperature discontinuity of amplitude[186)]

$$\frac{\delta T}{T} = 4\pi G\mu v\gamma(v)\,, \tag{7.15}$$

where v is the velocity of the string. Thus, the distinctive signature for cosmic strings in the microwave sky are line discontinuities in T of the above magnitude.

Figure 46: Sketch of the Kaiser-Stebbins effect by which cosmic strings produce linear discontinuities in the CMB. Photons γ passing on different sides of a moving string S (velocity v) towards the observer $\mathcal{O}$ receive a relative Doppler shift due to the conical nature of space perpendicular to the string (deficit angle α).

Given ideal maps of the CMB sky it would be easy to detect strings. However, real experiments have finite beam width. Taking into account averaging over a scale corresponding to the beam width will smear out the discontinuity, and it turns out to be surprisingly hard to distinguish the predictions of the cosmic string model from that of inflation-based theories using quantitative statistics which are easy to evaluate analytically, such as the kurtosis of the spatial gradient map of the CMB[187].

Textures produce a distribution of hot and cold spots on the CMB sky with typical size of several degrees[188]. This signature is much easier to see in CMB maps. The mechanism which produces these hot and cold spots in the CMB is illustrated in Fig. 47.

Figure 47: Space-time diagram of a collapsing texture. The unwinding occurs at the point TX. The shaded areas correspond to overdense regions. Photons like γ_1 are redshifted, those like γ_2 are blueshifted.

Photons arriving at the observer having passed through a texture as in the case of the ray γ_1 in Fig. 47 will be redshifted relative to the average photons since they have to climb out of a potential well, whereas those in orientation γ_2 will be blueshifted since they fall into a potential well. Taking into account reionization produced by texture collapse gives an amplitude of $\delta T/T$ of[189,190]

$$\frac{\delta T}{T} \sim 0.06 \times 16\pi G\eta^2 \,. \tag{7.16}$$

A number of about ten hot and cold spots of angular scale 10° is predicted by the texture model.

Theories of structure formation can now be normalized from CMB anisotropy data and from large-scale structure considerations. An inflationary model with CDM yields agreement between these two normalizations provided[191]

$$b \simeq 1 \,, \tag{7.17}$$

where b is the bias factor determining the ratio of fractional mass to light pertur-

bations on a scale of $8h^{-1}$ Mpc.

$$\frac{\delta L}{L}\Big|_{8h^{-1}\mathrm{Mpc}} = b\,\frac{\delta M}{M}\Big|_{8h^{-1}\mathrm{Mpc}}\,. \tag{7.18}$$

However, agreement between galaxy and cluster correlation properties seem to require[192)]

$$b \sim 2\,. \tag{7.19}$$

Normalizations of the texture model from large-scale structure and CMB observations[189,190,193)] require a bias

$$b \sim 3\,, \tag{7.20}$$

whereas for cosmic strings the two normalizations agree well. Based both on numerical simulations and analytical calculations, a normalization of the cosmic string model from the COBE CMB anisotropy data gives[194,195)]

$$G\mu = (1.3 \pm 0.5)10^{-6}\,. \tag{7.21}$$

7.3. Experimental Results

Over the past couple of years there has been a spectacular breakthrough on the observational front. The DMR experiment on the COBE satellite[196)] has produced a temperature map of the entire sky with beam width of 7°, which shows a clear detection of CMB anisotropies. Independent confirmation has come from two 5° experiments, FIRAS[197)] which has mapped 1/4 of the sky, and the Tenerife experiment[198)] which surveyed a strip of 70° length in right ascension at a declination 40°. The FIRAS data cross correlate very well with the COBE results, and there is even good agreement in the location of a pronounced feature in the Tenerife map with a that of a comparable feature in the two-year COBE maps.

In addition, there are many small angular scale experiments which have detected anisotropies. A partial list of observational results is given in Table 2 . In this table, "Angular Scale" denotes the beam width, the "results for $\delta T/T$" stands for the variance of δT computed from the CMB maps, "cover" indicates the area of the sky mapped. MAX 1 and MAX 2 denote two separate MAX measurements of $\delta T/T$, one in a region of the sky μ Peg, the second near GUM. OVRO 1 is the first Owens Valley experiment, a measurement near the North Galactic Cap, the second is a ring survey. The large numer of anisotropy experiments which have announced detections of temperature fluctuations since April 1992 indicates the rapid progress in this field.

To a first approximation, the present experimental results are in agreement with the predictions of a scale invariant spectrum of density perturbations. A popular way to show the results is to expand $T(\underline{\mathrm{n}})$ in spherical harmonics

$$T(\underline{\mathrm{n}}) = \sum_l \sum_{m=-l}^{l} a_{lm} Y_{lm}(\underline{\mathrm{n}}), \tag{7.22}$$

where $\underline{\mathrm{n}}$ is a unit vector on the sky, and to calculate the temperature correlation function

$$< T(\underline{\mathrm{n}}_1) T(\underline{\mathrm{n}}_2) >= \frac{1}{4\pi} \sum_l (2l+1) C_l P_l(\underline{\mathrm{n}}_1 \cdot \underline{\mathrm{n}}_2) \tag{7.23}$$

where

$$< a^*_{lm}\, a_{l'm'} >= C_l \delta_{ll'} \delta_{mm'}\, . \tag{7.24}$$

For a power spectrum of density perturbations

$$P(k) \sim k^n \tag{7.25}$$

the prediction for the Sachs-Wolfe contribution to δT is

$$l^2 C_l \sim l^{n-1} \tag{7.26}$$

on scales larger than the Hubble radius at t_{rec} (*i.e.*, for small values of l).

A direct comparison between theory and experiment is complicated by two effects: the Doppler contribution to $\delta T/T$ creates a peak in the $l^2 C_l$ curve at values of l which correspond to wavelengths comparable to the Hubble radius at t_{rec}, whose amplitude depends strongly on the ionization history of the Universe. Reionization also leads to a decrease in C_l for large l.

The COBE results combined with Tenerife observations favor[198] a value of n larger than what is predicted by simple inflationary models. However, the error bars are large and the difference is not (yet) statistically significant. At present there is the intriguing puzzle as to why the signal of certain small scale experiments is larger than the upper limit of other observations at the same angular scale elsewhere in the sky. A search for possible non-Gaussian features in the CMB sky will have high priority in the next years.

8. Modern Cosmology and Planck Scale Physics

8.1. Introduction

Through its implications for very early Universe cosmology, Planck scale physics (and specifically string theory) might well have directly observable consequences for the physical world. The aim of this chapter is to explore some possibilities of how this may occur.

As was explained in Chapter 5, standard particle physics models do not yield a convincing realization of inflation since in this context, inflation requires a fundamental scalar field with a reasonably flat potential (in order to have inflation) and with very small coupling constants (in order that quantum fluctuations present during inflation do not lead to CMB temperature anisotropies in excess of those recently detected. Such potentials are not generic in particle physics models.

TABLE 2: CMB Anisotropy Results

Experiment	Angular Scale	Result for $\frac{\delta T}{T}$	Cover	Location
COBE-DMR[196]	7°	1.1 ± 0.2	4π	space
Tenerife[198]	5.6°	1.7 ± 0.4	350 deg^2	ground
FIRS[197]	4°	1 - 3	π	balloon
SK93[199]	1.45°	1.4 ± 0.5		ground
SP91[200]	1.4°	1.1 ± 0.5	13.8 deg^2	ground
ARGO[201]	1°	2.2 ± 0.8		balloon
Python[202]	0.75°	3	8 deg^2	ground
MAX 1[203]	0.5°	< 3		balloon
MAX 2[204]	0.5°	4.9 ± 0.8	6 deg^2	balloon
MSAM[205]	0.47°	1.6 ± 0.4	6 deg^2	balloon
White Dish[206]	0.2°	< 2.3		ground
OVRO 1[207]	$1.8'$	< 1.9	0.03 deg^2	ground
OVRO 2[208]	$1.8'$	3.4 ± 1.1	0.1 deg^2	ground

The first challenge from cosmologists to Planck scale physics is therefore to provide a generic mechanism for inflation. It may be that Planck scale physics predicts the type of scalar field potentials for which successful inflation results.

Another possibility is that Planck scale physics leads to a realization of inflation which does not involve scalar fields. A possible scenario for this is suggested in Section 2. Finally, it may be that Planck scale physics leads to a solution of the homogeneity and flatness problems which does not require inflation.

Standard and modern cosmology are plagued by an internal inconsistency. They predict that the Universe started at a "Big Bang" singularity with infinite curvature and matter temperature. However, it is known that the physics on which the standard cosmological model is built must break down at very high temperature and curvature. Therefore, the second challenge for Planck scale physics is to find a solution to the singularity problem. Two very different scenarios in which this may happen are suggested in Sections 2 and 3.

Finally, Planck scale physics (string theory as a concrete example) allows us to ask questions about the physical world which cannot be posed in standard physics. For example, is there a dynamical mechanism which singles out a Universe in which three space and one time dimensions are observable? One mechanism in the context of string theory will be reviewed in Section 3.

I will review two very different approaches to Planck scale cosmology. The first is an attempt to incorporate Planck scale effects on the space-time structure by writing down an effective action for the space-time metric. It will be shown that a class of effective actions exists whose solutions have a less singular structure. More specifically, all homogeneous and isotropic solutions are nonsingular (see Section 2).

In Section 3, I will summarize some aspects of string cosmology and indicate how in the context of string theory the cosmological singularities can be avoided. A dynamical mechanism which explains why at most three-spatial dimensions are large (and thus observable) is suggested.

8.2. A Nonsingular Universe

Motivation

Planck scale physics will generate corrections to the Einstein action which determines the dynamics of the space-time metric $g_{\mu\nu}$. This can be seen by considering the effective action obtained by integrating out quantum matter fields in the presence of a dynamical metric, by calculating first order perturbative quantum gravity effects, or by studying the low energy effective action of a Planck scale unified theory such as string theory.

The question we wish to address in this section is whether it is possible to construct a class of effective actions for gravity which have improved singularity properties and which predict inflation, with the constraint that they give the correct low curvature limit.

What follows is a summary of recent work[41,218,219] in which we have constructed an effective action for gravity in which all solutions with sufficient symmetry are nonsingular. The theory is a higher derivative modification of the Einstein action, and is obtained by a constructive procedure well motivated in analogy with the analysis of point particle motion in special relativity. The resulting theory is asymptotically free in a sense which will be specified below.

A possible objection to our approach is that near a singularity quantum effects will be important and therefore a classical analysis is doomed to fail. This argument is correct in the usual picture in which at high curvatures there are large fluctuations and space-time becomes more like a "quantum foam." However, in our theory, at high curvature space-time becomes highly regular and thus a classical analysis of space-time is self-consistent. The property of asymptotic freedom is essential in order to reach this conclusion.

Our aim is to construct a theory with the property that the metric $g_{\mu\nu}$ approaches the de Sitter metric $g^{DS}_{\mu\nu}$, a metric with maximal symmetry which admits a geodesically complete and nonsingular extension, as the curvature R approaches

the Planck value R_{pl}. Here, R stands for any curvature invariant. Naturally, from our classical considerations, R_{pl} is a free parameter. However, if our theory is connected with Planck scale physics, we expect R_{pl} to be set by the Planck scale.

Figure 48: Penrose diagrams for collapsing Universe (left) and black hole (right) in Einstein's theory (top) and in the nonsingular Universe (bottom). C, E, DS and H stand for contracting phase, expanding phase, de Sitter phase and horizon, respectively, and wavy lines indicate singularities.

If successful, the above construction will have some very appealing consequences. Consider, for example, a collapsing spatially homogeneous Universe. According to Einstein's theory, this Universe will collapse in finite proper time to a final "big crunch" singularity (top left Penrose diagram of Figure 48). In our theory, however, the Universe will approach a de Sitter model as the curvature increases. If the Universe is closed, there will be a de Sitter bounce followed by re-expansion (bottom left Penrose diagram in Figure 48). Similarly, in our theory spherically symmetric vacuum solutions would be nonsingular, i.e., black holes would have no singularities in their centers. The structure of a large black hole would be unchanged compared to what is predicted by Einstein's theory (top right, Figure 48) outside and even slightly inside the horizon, since all curvature invariants are small in those regions. However, for $r \to 0$ (where r is the radial Schwarzschild coordinate), the solution changes and approaches a de Sitter solution (bottom right, Figure 48). This would have interesting consequences for the black hole information loss problem.

To motivate our effective action construction, we turn to a well known analogy, point particle motion in the theory of special relativity.

An Analogy

The transition from the Newtonian theory of point particle motion to the special relativistic theory transforms a theory with no bound on the velocity into one in which there is a limiting velocity, the speed of light c (in the following we

use units in which $\hbar = c = 1$). This transition can be obtained[41)] by starting with the action of a point particle with world line $x(t)$:

$$S_{\text{old}} = \int dt \frac{1}{2}\dot{x}^2 \,, \tag{8.1}$$

and adding[220)] a Lagrange multiplier which couples to $\dot{x}^2$, the quantity to be made finite, and which has a potential $V(\varphi)$:

$$S_{\text{new}} = \int dt \left[\frac{1}{2}\dot{x}^2 + \varphi\dot{x}^2 - V(\varphi)\right] . \tag{8.2}$$

From the constraint equation

$$\dot{x}^2 = \frac{\partial V}{\partial \varphi} \,, \tag{8.3}$$

it follows that $\dot{x}^2$ is limited provided $V(\varphi)$ increases no faster than linearly in φ for large $|\varphi|$. The small φ asymptotics of $V(\varphi)$ is determined by demanding that at low velocities the correct Newtonian limit results:

$$\begin{aligned} V(\varphi) &\sim \varphi^2 \text{ as } |\varphi| \to 0 \,, \\ V(\varphi) &\sim \varphi \text{ as } |\varphi| \to \infty \,. \end{aligned} \tag{8.4}$$

Choosing the simple interpolating potential

$$V(\varphi) = \frac{2\varphi^2}{1 + 2\varphi} \,, \tag{8.5}$$

the Lagrange multiplier can be integrated out, resulting in the well-known action

$$S_{\text{new}} = \frac{1}{2} \int dt \sqrt{1 - \dot{x}^2} \tag{8.6}$$

for point particle motion in special relativity.

Construction

Our procedure for obtaining a nonsingular Universe theory[41] is based on generalizing the above Lagrange multiplier construction to gravity. Starting from the Einstein action, we can introduce a Lagrange multiplier φ_1 coupled to the Ricci scalar R to obtain a theory with limited R:

$$S = \int d^4x\sqrt{-g}(R + \varphi_1 R + V_1(\varphi_1)) \,, \tag{8.7}$$

where the potential $V_1(\varphi_1)$ satisfies the asymptotic conditions (8.4).

However, this action is insufficient to obtain a nonsingular gravity theory. For example, singular solutions of the Einstein equations with $R = 0$ are not effected at all. The minimal requirements for a nonsingular theory is that all curvature invariants remain bounded and the space-time manifold is geodesically complete. Implementing the limiting curvature hypothesis[221], these conditions can be reduced to more manageable ones. First, we choose one curvature invariant $I_1(g_{\mu\nu})$ and demand that it be explicitely bounded, i.e., $|I_1| < I_1^{pl}$, where I_1^{pl} is the Planck scale value of I_1. In a second step, we demand that as $I_1(g_{\mu\nu})$ approaches I_1^{pl}, the metric $g_{\mu\nu}$ approach the de Sitter metric $g_{\mu\nu}^{DS}$, a definite nonsingular metric with maximal symmetry. In this case, all curvature invariants are automatically bounded (they approach their de Sitter values), and the space-time can be extended to be geodesically complete.

Our approach is to implement the second step of the above procedure by another Lagrange multiplier construction[41]. We look for a curvature invariant $I_2(g_{\mu\nu})$ with the property that

$$I_2(g_{\mu\nu}) = 0 \;\; \Leftrightarrow \;\; g_{\mu\nu} = g_{\mu\nu}^{DS} \,, \tag{8.8}$$

introduce a second Lagrange multiplier field φ_2 which couples to I_2 and choose a

potential $V_2(\varphi_2)$ which forces I_2 to zero at large $|\varphi_2|$:

$$S = \int d^4x\sqrt{-g}[R + \varphi_1 I_1 + V_1(\varphi_1) + \varphi_2 I_2 + V_2(\varphi_2)]\,, \tag{8.9}$$

with asymptotic conditions (8.4) for $V_1(\varphi_1)$ and conditions

$$\begin{aligned} &V_2(\varphi_2) \sim \text{const as } |\varphi_2| \to \infty \\ &V_2(\varphi_2) \sim \varphi_2^2 \text{ as } |\varphi_2| \to 0\,, \end{aligned} \tag{8.10}$$

for $V_2(\varphi_2)$. The first constraint forces I_2 to zero, the second is required in order to obtain the correct low curvature limit.

These general conditions are reasonable, but not sufficient in order to obtain a nonsingular theory. It must still be shown that all solutions are well behaved, i.e., that they asymptotically reach the regions $|\varphi_2| \to \infty$ of phase space (or that they can be controlled in some other way). This must be done for a specific realization of the above general construction.

Specific Model

At the moment we are only able to find an invariant I_2 which singles out de Sitter space by demanding $I_2 = 0$ provided we assume that the metric has special symmetries. The choice

$$I_2 = (4R_{\mu\nu}R^{\mu\nu} - R^2 + C^2)^{1/2}\,, \tag{8.11}$$

singles out the de Sitter metric among all homogeneous and isotropic metrics (in which case adding C^2, the Weyl tensor square, is superfluous), all homogeneous and anisotropic metrics, and all radially symmetric metrics.

We choose the action[41)]

$$S = \int d^4x\sqrt{-g}\left[R + \varphi_1 R - (\varphi_2 + \frac{3}{\sqrt{2}}\varphi_1)I_2^{1/2} + V_1(\varphi_1) + V_2(\varphi_2)\right] \tag{8.12}$$

with

$$V_1(\varphi_1) = 12\, H_0^2 \frac{\varphi_1^2}{1+\varphi_1}\left(1 - \frac{\ln(1+\varphi_1)}{1+\varphi_1}\right) \tag{8.13}$$

$$V_2(\varphi_2) = -2\sqrt{3}\, H_0^2\, \frac{\varphi_2^2}{1+\varphi_2^2}\,. \tag{8.14}$$

The general equations of motion resulting from this action are quite messy. However, when restricted to homogeneous and isotropic metrics of the form

$$ds^2 = dt^2 - a(t)^2(dx^2 + dy^2 + dz^2)\,, \tag{8.15}$$

the equations are fairly simple. With $H = \dot{a}/a$, the two φ_1 and φ_2 constraint equations are

$$H^2 = \frac{1}{12} V_1' \tag{8.16}$$

$$\dot{H} = -\frac{1}{2\sqrt{3}} V_2'\,, \tag{8.17}$$

and the dynamical g_{00} equation becomes

$$3(1-2\varphi_1)H^2 + \frac{1}{2}(V_1 + V_2) = \sqrt{3} H(\dot{\varphi}_2 + 3H\varphi_2)\,. \tag{8.18}$$

The phase space of all vacuum configurations is the half plane $\{(\varphi_1 \geq 0,\, \varphi_2)\}$. Equations (8.16) and (8.17) can be used to express H and $\dot{H}$ in terms of φ_1 and φ_2. The remaining dynamical equation (8.18) can then be recast as

$$\frac{d\varphi_2}{d\varphi_1} = -\frac{V_1''}{4V_2'}\left[-\sqrt{3}\varphi_2 + (1-2\varphi_1) - \frac{2}{V_1'}(V_1 + V_2)\right]\,. \tag{8.19}$$

The solutions can be studied analytically in the asymptotic regions and numerically throughout the entire phase space.

The resulting phase diagram of vacuum solutions is sketched in Fig. 49 (for numerical results, see the second article in Ref. 41). The point $(\varphi_1, \varphi_2) = (0, 0)$ corresponds to Minkowski space-time M^4, the regions $|\varphi_2| \to \infty$ to de Sitter space. As shown, all solutions either are periodic about M^4 or else they asymptotically approach de Sitter space. Hence, all solutions are nonsingular. This conclusion remains unchanged if we add spatial curvature to the model.

Figure 49: Phase diagram of the homogeneous and isotropic solutions of the nonsingular Universe. The asymptotic regions are labelled by A, B, C and D, flow lines are indicated by arrows.

One of the most interesting properties of our theory is asymptotic freedom[41)], i.e., the coupling between matter and gravity goes to zero at high curvatures. It is easy to add matter (e.g., dust or radiation) to our model by taking the combined action

$$S = S_g + S_m \,, \tag{8.20}$$

where S_g is the gravity action previously discussed, and S_m is the usual matter action in an external background space-time metric.

We find[41)] that in the asymptotic de Sitter regions, the trajectories of the solutions in the (φ_1, φ_2) plane are unchanged by adding matter. This applies, for example, in a phase of de Sitter contraction when the matter energy density is increasing exponentially but does not affect the metric. The physical reason for asymptotic freedom is obvious: in the asymptotic regions of phase space, the space-time curvature approaches its maximal value and thus cannot be changed even by adding an arbitrary high matter energy density.

Naturally, the phase space trajectories near $(\varphi_1, \varphi_2) = (0, 0)$ are strongly effected by adding matter. In particular, M^4 ceases to be a stable fixed point of the evolution equations.

Connection with Dilaton Gravity

The low energy effective actions for the space-time metric in 4 dimensions which come from string theory are only known perturbatively. They contain higher derivative terms, but not if the exact same form as the ones used in our construction. The connection between our limiting curvature construction and string theory-motivated effective actions is more apparent in two space-time dimensions[218,219].

The most general renormalizable Lagrangian for string-induced dilaton gravity is

$$\mathcal{L} = \sqrt{-g}[D(\varphi)R + G(\varphi)(\nabla\varphi)^2 + H(\varphi)]\,, \tag{8.21}$$

where $\varphi(x,t)$ is the dilaton. In two space-time dimensions, the kinetic term for φ can be eliminated, resulting in a Lagrangian (in terms of rescaled fields) of the form

$$\mathcal{L} = \sqrt{-g}[D(\varphi)R + V(\varphi)]\,. \tag{8.22}$$

We can now apply the limiting curvature construction to find classes of potentials for which the theory has nonsingular black hole[218] and cosmological[219] solutions. In the following, we discuss the nonsingular two-dimensional black hole.

To simplify the algebra, the dilaton is redefined such that

$$D(\varphi) = \frac{1}{\varphi}\,. \tag{8.23}$$

The most general static metric can be written as

$$ds^2 = f(r)dt^2 - g(r)dr^2 \tag{8.24}$$

and the gauge choice

$$g(r) = f(r)^{-1} \tag{8.25}$$

is always possible. The variational equations are

$$f' = -V(\varphi)\frac{\varphi^2}{\varphi'}\,, \tag{8.26}$$

$$\left(\frac{\varphi'}{\varphi^2}\right)' = 0 \tag{8.27}$$

and

$$\varphi^{-2}R = \frac{\partial V}{\partial \varphi}\,, \tag{8.28}$$

where a prime denotes the derivative with respect to r.

Equation (8.27) can be integrated to find (after rescaling r)

$$\varphi = \frac{1}{Ar}\,. \tag{8.29}$$

To give the correct large r behavior for the metric, we need to impose that

$$f(r) \to 1 - \frac{2m}{r} \quad \text{as } r \to \infty\,. \tag{8.30}$$

From (8.26) this leads to the asymptotic condition

$$V(\varphi) \to 2mA^3\varphi^2 \quad \text{as } \varphi \to 0\,. \tag{8.31}$$

The limiting curvature hypothesis requires that R be bounded as $\varphi \to \infty$. From (8.28) this implies

$$V(\varphi) \to \frac{2}{\ell^2\varphi} \quad \text{as } \varphi \to \infty\,, \tag{8.32}$$

where ℓ is a constant which determines the limiting curvature. As an interpolating potential we can choose

$$V(\varphi) = \frac{2mA^3\varphi^2}{1 + mA^3\ell^2\varphi^3}\,, \tag{8.33}$$

which allows (8.26) to be integrated explicitly[218)] to obtain $f(r)$.

The resulting metric coefficient $f(r)$ describes a nonsingular black hole with a single horizon at $r \simeq 2m$. The metric is indistinguishable from the usual Schwarzschild metric until far inside of the horizon, where our $f(r)$ remains regular and obtains vanishing derivative at $r = 0$, which allows for a geodesically complete extension of the manifold.

Discussion

We have shown that a class of higher derivative extensions of the Einstein theory exist for which many interesting solutions are nonsingular. This class of models is very special. Most higher derivative theories of gravity have, in fact, much worse singularity properties than the Einstein theory. What is special about this class of theories is that they are obtained using a well motivated Lagrange multiplier construction which implements the limiting curvature hypothesis. We have shown that

i) all homogeneous and isotropic solutions are nonsingular[41)]

ii) the two-dimensional black holes are nonsingular[218)]

iii) nonsingular two-dimensional cosmologies exist[219)].

We also have evidence that four-dimensional black holes and anisotropic homogeneous cosmologies are nonsingular[222)].

By construction, all solutions are de Sitter at high curvature. Thus, the theories automatically have a period of inflation (driven by the gravity sector in analogy to Starobinsky inflation[39)]) in the early Universe.

A very important property of our theories is asymptotic freedom. This means that the coupling between matter and gravity goes to zero at high curvature, and might lead to an automatic suppression mechanism for scalar fluctuations.

In two space-time dimensions, there is a close connection between dilaton gravity and our construction. In four dimensions, the connection between fundamental physics and our class of effective actions remains to be explored. In particular, it

would be nice to investigate the connection between our limiting curvature construction and the 'pre-big-bang cosmology' scenario proposed on the basis of dilaton gravity in Ref. 119.

8.3. Aspects of String Cosmology

Motivation

In the previous section we studied effective actions for the space-time metric which might arise in the intermediate energy regime of a fundamental theory such as string theory. However, it is also of interest to explore the predictions of string theory which depend specifically on the "stringy" aspects of the theory and which are lost in any field theory limit. It is to a description of a few of the string-specific cosmological aspects to which we turn in this section.

Implications of Target Space Duality

Target space duality[223] is a symmetry specific to string theory. As a simple example, consider a superstring background in which all spatial dimensions are toroidally compactified with equal radii. Let R denote the radius of the torus.

The spectrum of string states is spanned by oscillatory modes which have energies independent of R, by momentum modes whose energies E_n (with integer n) are

$$E_n = \frac{n}{R}\,, \tag{8.34}$$

and by winding modes with energies E'_m (m integer)

$$E'_m = mR\,. \tag{8.35}$$

Target space duality is a symmetry between two superstring theories, one on a background with radius R, the other on a background of radius $1/R$, under which winding and momentum modes are interchanged.

Target space duality has interesting consequences for string cosmology[224]. Consider a background with adiabatically changing $R(t)$. While $R(t) \gg 1$, most of the energy in thermal equilibrium resides in the momentum modes. The position eigenstates $|x>$ are defined as in quantum field theory in terms of the Fourier transform of the momentum eigenstates $|p>$

$$|x>=\sum_p e^{ix\cdot p}|p> \ . \tag{8.36}$$

However, for $R(t) \ll 1$, most of the energy flows into winding modes, and it takes much less energy to measure the "dual distance" $|\tilde{x}>$ than $|x>$, where

$$|\tilde{x}>=\sum_w e^{i\tilde{x}\cdot w}|w> \tag{8.37}$$

is defined in terms of the winding modes $|w>$.

We conclude that target space duality in string theory leads to a minimum physical length in string cosmology. As $R(t)$ decreases below 1, the measured length starts to increase again. This could lead to a bouncing or oscillating cosmology[224].

It is well known that for strings in thermal equilibrium there is a maximal temperature, the Hagedorn temperature[225]. Target space duality implies that in thermal equilibrium the temperature in an adiabatically varying string background begins to decrease once $R(t)$ falls below 1:

$$T\left(\frac{1}{R}\right) = T(R)\,. \tag{8.38}$$

Thus, the $T(R)$ curve in string cosmology is nonsingular and very different from its behavior in standard cosmology. For further discussions of the thermodynamics of strings see, e.g., Refs. 226 and 227 and references therein.

Strings and Space-Time Dimensionality

Computations[224)] using the microcanonical ensemble show that for all spatial directions compactified at large total energy E, the entropy S is proportional to E:

$$S = \beta_H E \,, \tag{8.39}$$

with β_H denoting the inverse of the Hagedorn temperature T_H. Thus, the $E(R)$ curve in string cosmology is very different from the corresponding curve in standard cosmology.

For large $R \gg 1$, most of the energy in a gas of strings in thermal equilibrium will flow into momentum modes, and the thermodynamics will approach that of an ideal gas of radiation for which

$$E(R) \sim \frac{1}{R} \,. \tag{8.40}$$

By duality, for small R

$$E(R) \sim R \,. \tag{8.41}$$

If, however, for some reason the string gas falls out of equilibrium, the $E(R)$ curve will look very different. Starting at $R = 1$ with a temperature approximately equal to T_H, a large fraction of the energy will reside in winding modes. If these winding modes cannot annihilate, thermal equilibrium will be lost, and the energy in winding modes will increase linearly in R, and thus for large R:

$$E(R) \sim R \,. \tag{8.42}$$

Newtonian intuition tells us that out of equilibrium winding modes with an energy relation (8.42) will prevent the background space from expanding[224]. The

equation of state corresponding to a gas of straight strings is

$$p = -\frac{1}{N}\rho \tag{8.43}$$

where p and ρ denote pressure and energy density, respectively, and N is the number of spatial dimensions. According to standard general relativity, an equation of state with negative pressure will lead to more rapid expansion of the background. It turns out that the Newtonian intuition is the correct one and that general relativity gives the wrong answer[228]. At high densities, the specific stringy effects – in particular target space duality – become crucial.

The Einstein action violates duality. In order to restore duality, it is necessary to include the dilaton in the effective action for the string background. The action for dilaton gravity is

$$S = \int d^{N+1}x\sqrt{-g}e^{-2\phi}[R + 4(D\phi)^2] \tag{8.44}$$

where ϕ is the dilaton. It is convenient to use new fields φ and λ defined by

$$a(t) = e^{\lambda t} \tag{8.45}$$

and

$$\varphi = 2\phi - N\lambda\,. \tag{8.46}$$

The action (8.44) has the duality symmetry

$$\lambda \to -\lambda,\ \varphi \to \varphi\,. \tag{8.47}$$

The variational equations of motion derived from (8.44) for a homogeneous

and isotropic model are[228,229)]

$$\begin{aligned}
\dot{\varphi}^2 &= e^{\varphi}E + N\dot{\lambda}^2 \\
\ddot{\lambda} - \dot{\varphi}\dot{\lambda} &= \frac{1}{2}e^{\varphi}P \\
\ddot{\varphi} &= \frac{1}{2}e^{\varphi}E + N\dot{\lambda}^2\,,
\end{aligned} \tag{8.48}$$

where P and E are total pressure and energy, respectively. For a winding mode-dominated equation of state (and neglecting friction terms) the equation of motion for $\lambda(t)$ becomes

$$\ddot{\lambda} = -\frac{1}{2N}e^{\varphi}E(\lambda)\,, \tag{8.49}$$

which corresponds to motion in a confining potential. Hence, winding modes prevent the background toroidal dimensions from expanding.

These considerations may be used to put forward the conjecture[224)] that string cosmology will single out three as the maximum number of spatial dimensions which can be large ($R \gg 1$ in Planck units). The argument proceeds as follows. Space can, starting from an initial state with $R \sim 1$ in all directions, only expand if thermal equilibrium is maintained, which in turn is only possible if the winding modes can annihilate. This can only happen in at most three spatial dimensions (in a higher number the probability for intersection of the world sheets of two strings is zero). In the critical dimension for strings, $N = 3$, the evolution of a string gas has been studied extensively in the context of the cosmic string theory (see Chapter 6). The winding modes do, indeed, annihilate, leaving behind a string network with about one winding mode passing through each Hubble volume. Thus, in string cosmology only three spatial dimensions will become large whereas the others will be confined to Planck size by winding modes.

8.4. Summary

Planck scale physics may have many observational consequences and may help cosmologists solve some of the deep puzzles concerning the origin of inflation, the absence of space-time singularities and the dimensionality of space-time.

A lot of work needs to be done before these issues are properly understood. I have outlined two ways to address some of these questions. The first investigation was based on classical physics and attempted to analyze what can be said about the origin of inflation and about singularities from an effective action approach to gravity. We constructed a class of higher derivative gravity actions without singular cosmological solutions (i.e., no singular homogeneous and isotropic solutions) and which automatically give rise to inflation.

The second approach was an exploration of some of the cosmological consequences of target space duality in string theory. A nonsingular cosmological scenario was proposed which might even explain why only three-spatial dimensions are large.

9. Conclusions

Modern cosmology has led to the development of several theories of structure formation, most prominently theories based on inflation, and topological defect models. These new theories are all based on the union between particle physics and general relativity. The models of structure formation obey the usual causality principle of relativistic physics.

All of the current theories of structure formation have their problems. Most importantly, they do not address the cosmological constant problem but rather, inasmuch as they make use of scalar matter fields, make the problem worse. The inflationary Universe scenario is still lacking a convincing realization. Present versions require very special scalar field potentials. Topological defect models,

on the other hand, do not explain why the Universe is nearly homogeneous and spatially flat (however, they are consistent with a low Ω Universe). In my opinion, we should regard our current theories as toy models with which we work and from which we learn, but which will eventually be replaced by improved and more convincing theories.

Nevertheless, our present theories are predictive. To a first approximation, they all predict a scale invariant spectrum of density perturbations and induced CMB anisotropies. Typically, the models contain one intrinsically free parameter (plus maybe a couple more parameters with which we can describe our ignorance of the detailed evolution of the models). The free parameter can be normalized from any one of several observables. It is remarkable that the different normalizations of the models are consistent (to a first approximation). This lets us entertain the hope that we are on the right track: structure formation proceeds via gravitational instability (Sections 4.3 and 4.4) with the seed perturbations being provided by a particle physics theory of the very early Universe.

There is already a wealth of observational data which is fit quite well by our present toy models. More and higher accuracy data is rapidly becoming available. The data concerns on one hand structure in the Universe gleamed from optical and infrared galaxy surveys, and on the other hand from the temperature map of the CMB sky.

With the wealth of data available and steady flow of new observational results, and given that many important questions remain unresolved, modern cosmology will remain an exciting area of research for the forseeable future.

The basic problems which are not addressed by our present theories of cosmology might be resolved by some as yet unknown unified theory of all forces. Some speculations along these lines were entertained in the last chapter of these lecture notes. It is of particular interest to investigate whether string theory leads to a more convincing realization of inflation, and whether there is a mechanism which predicts why our Universe consists of three large spatial dimensions.

Acknowledgments

I wish to thank Professor Mario Novello for inviting me to give these lectures in Angra, and all the organizers and participants for their wonderful hospitality and for their many stimulating questions. I am grateful to all of my research collaborators, on whose work I have freely drawn. Partial financial support for the preparation of this manuscript has been provided at Brown by the US Department of Energy under Grant DE-FG0291ER40688, Task A, and at UBC by the Canadian NSERC under Grant 580441.

REFERENCES

1. A. Linde, 'Particle Physics and Inflationary Cosmology' (Harwood, Chur, 1990).
2. S. Blau and A. Guth, 'Inflationary Cosmology,' in '300 Years of Gravitation' ed. by S. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1987).
3. K. Olive, *Phys. Rep.* **190**, 307 (1990).
4. T.W.G. Kibble, *Phys. Rep.* **67**, 183 (1980).
5. A. Vilenkin, *Phys. Rep.* **121**, 263 (1985).
6. N. Turok, 'Phase Transitions as the Origin of Large-Scale Structure,' in 'Particles, Strings and Supernovae' (TASI-88) ed. by A. Jevicki and C.-I. Tan (World Scientific, Singapore, 1989).
7. A. Vilenkin and E.P.S. Shellard, 'Strings and Other Topological Defects' (Cambridge Univ. Press, Cambridge, 1994).
8. R. Brandenberger, *Rev. Mod. Phys.* **57**, 1 (1985).
9. R. Brandenberger, "Modern Cosmology and Structure Formation", in 'CP Violation and the Limits of the Standard Model (TASI94)', ed. J. Donoghue (World Scientific, Singapore, 1995).

10. R. Brandenberger, in 'Physics of the Early Universe,' proc. of the 1989 Scottish Univ. Summer School in Physics, ed. by J. Peacock, A. Heavens and A. Davies (SUSSP Publ., Edinburgh, 1990);
R. Brandenberger, in '1991 Summer School in High Energy Physics and Cosmology', eds. E. Gava et al. (World Scientific, Singapore, 1992);
R. Brandenberger, 'Lectures on Modern Cosmology and Structure Formation', in 'Particles and Fields', ed. by O. Eboli and V. Ribelles (World Scientific, Singapore 1994).

11. V. Mukhanov, H. Feldman and R. Brandenberger, *Phys. Rep.* **215**, 203 (1992).

12. T.W.B. Kibble, *J. Phys.* **A9**, 1387 (1976).

13. E. Milne, *Zeits. f. Astrophys.* **6**, 1 (1933).

14. V. de Lapparent, M. Geller and J. Huchra, *Ap. J. (Lett)* **302**, L1 (1986).

15. S. Shechtman, P. Schechter, A. Oemler, D. Tucker, R. Kirshner and H. Lin, Harvard-Smithsonian preprint CFA 3385 (1992), to appear in 'Clusters and Superclusters of Galaxies', ed. by A. Fabian (Kluwer, Dordrecht, 1993);
S. Shechtman et al., 'The Las Campanas Fiber-Optic Redshift Survey', CFA preprint (1994), to be publ. in the proc. of the 35th Herstmonceaux Conference 'Wide Field Spectroscopy and the Distant Universe'.

16. R. Partridge, *Rep. Prog. Phys.* **51**, 647 (1988).

17. see e.g., S. Weinberg, 'Gravitation and Cosmology' (Wiley, New York, 1972);
Ya.B. Zel'dovich and I. Novikov, 'The Structure and Evolution of the Universe' (Univ. of Chicago Press, Chicago, 1983).

18. E. Hubble, *Proc. Nat. Acad. Sci.* **15**, 168 (1927).

19. J. Mould et al., *Ap. J.* **383**, 467 (1991).

20. R. Alpher and R. Herman, *Rev. Mod. Phys.* **22**, 153 (1950);
G. Gamov, *Phys. Rev.* **70**, 572 (1946).

21. R. Dicke, P.J.E. Peebles, P. Roll and D. Wilkinson, *Ap. J.* **142**, 414 (1965).

22. A. Penzias and R. Wilson, *Ap. J.* **142**, 419 (1965).

23. J. Mather et al., *Ap. J. (Lett.)* **354**, L37 (1990).

24. H. Gush, M. Halpern and E. Wishnow, *Phys. Rev. Lett.* **65**, 937 (1990).

25. R. Alpher, H. Bethe and G. Gamov, *Phys. Rev.* **73**, 803 (1948);
R. Alpher and R. Herman, *Nature* **162**, 774 (1948).

26. For an excellent introduction see S. Weinberg, 'The First Three Minutes' (Basic Books, New York, 1988).

27. T. Padmanabhan, 'Structure Formation in the Universe' (Cambridge Univ. Press, Cambridge, 1993), Chap. 3 and refs. therein.

28. H. Arp, G. Burbidge, F. Hoyle, J. Narlikar and N. Vickramasinghe, *Nature* **346**, 807 (1990).

29. P.J.E. Peebles, D. Schramm, E. Turner and R. Kron, *Nature* **352**, 769 (1991).

30. A. Guth, *Phys. Rev.* **D23**, 347 (1981).

31. P.J.E. Peebles, 'Principles of Physical Cosmology' (Princeton Univ. Press, Princeton, 1993).

32. J. Ostriker and L. Cowie, *Ap. J. (Lett.)* **243**, L127 (1981).

33. S. Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989);
S. Carroll, W. Press and E. Turner, *Ann. Rev. Astron. Astrophys.* **30**, 499 (1992).

34. D. Kazanas, *Ap. J.* **241**, L59 (1980).

35. W. Press, *Phys. Scr.* **21**, 702 (1980).

36. G. Chibisov and V. Mukhanov, 'Galaxy Formation and Phonons,' Lebedev Physical Institute Preprint No. 162 (1980);
G. Chibisov and V. Mukhanov, *Mon. Not. R. Astron. Soc.* **200**, 535 (1982).

37. V. Lukash, *Pis'ma Zh. Eksp. Teor. Fiz.* **31**, 631 (1980).

38. K. Sato, *Mon. Not. R. Astron. Soc.* **195**, 467 (1981).

39. A. Starobinsky, *Phys. Lett.* **91B**, 99 (1980).

40. M. Mijic, M. Morris and W.-M. Suen, *Phys. Rev.* **D34**, 2934 (1986).

41. V. Mukhanov and R. Brandenberger, *Phys. Rev. Lett.* **68**, 1969 (1992); R. Brandenberger, V. Mukhanov and A. Sornborger, *Phys. Rev.* **D48**, 1629 (1993).

42. J. Ye and R. Brandenberger, *Nucl. Phys.* **B346**, 149 (1990).

43. H. Nielsen and P. Olesen, *Nucl. Phys.* **B61**, 45 (1973).

44. R. Davis, *Phys. Rev.* **D35**, 3705 (1987).

45. N. Turok, *Phys. Rev. Lett.* **63**, 2625 (1989).

46. see e.g., G. Ross, Grand Unified Theories (Benjamin, Reading, 1985).

47. Ya.B. Zel'dovich, *Mon. Not. R. astron. Soc.* **192**, 663 (1980).

48. A. Vilenkin, *Phys. Rev. Lett.* **46**, 1169 (1981).

49. J. Primack, D. Seckel and B. Sadoulet, *Ann. Rev. Nucl. Part. Sci.* **38**, 751 (1988).

50. T. van Albada and R. Sancisi, *Phil. Trans. R. Soc. London*, **A320**, 447 (1986).

51. V. Trimble, *Ann. Rev. Astr. Astrophys.* **25**, 423 (1987).

52. E. Bertschinger and A. Dekel, *Ap. J.* **336**, L5 (1989).

53. M. Strauss, M. Davis, A. Yahil and J. Huchra, *Ap. J.* **385**, 421 (1992).

54. S. White, C. Frenk and M. Davis, *Ap. J. (Lett.)* **274**, L1 (1993).

55. J. Bond, G. Efstathiou and J. Silk, *Phys. Rev. Lett.* **45**, 1980 (1980); J. Bond and A. Szalay, *Ap. J.* **274**, 443 (1983).

56. G. Blumenthal, S. Faber, J. Primack and M. Rees, *Nature* **311**, 517 (1984); M. Davis, G. Efstathiou, C. Frenk and S. White, *Ap. J.* **292**, 371 (1985).

57. N. Turok and R. Brandenberger, *Phys. Rev.* **D33**, 2175 (1986);
A. Stebbins, *Ap. J. (Lett.)* **303**, L21 (1986);
H. Sato, *Prog. Theor. Phys.* **75**, 1342 (1986).

58. T. Vachaspati, *Phys. Rev. Lett.* **57**, 1655 (1986).

59. A. Stebbins, S. Veeraraghavan, R. Brandenberger, J. Silk and N. Turok, *Ap. J.* **322**, 1 (1987).

60. N. Turok, *Phys. Scripta* **T36**, 135 (1991).

61. R. Brandenberger, N. Kaiser, D. Schramm and N. Turok, *Phys. Rev. Lett.* **59**, 2371 (1987).

62. R. Brandenberger, N. Kaiser and N. Turok, *Phys. Rev.* **D36**, 2242 (1987).

63. R. Brandenberger, L. Perivolaropoulos and A. Stebbins, *Int. J. of Mod. Phys.* **A5**, 1633 (1990);
L. Perivolarapoulos, R. Brandenberger and A. Stebbins, *Phys. Rev.* **D41**, 1764 (1990);
R. Brandenberger, *Phys. Scripta* **T36**, 114 (1991).

64. M. Davis and J. Huchra, *Ap. J.* **254**, 437 (1982).

65. N. Bahcall and R. Soneira, *Ap. J.* **270**, 20 (1983);
A. Klypin and A. Kopylov, *Sov. Astr. Lett.* **9**, 41 (1983).

66. M. Strauss et al., *Ap. J.* **385**, 421 (1992).

67. A. Yahil et al., *Ap. J.* **372**, 380 (1991).

68. G. Abell, *Ap. J. Suppl.* **3**, 211 (1958).

69. G. Daulton et al., 'The two-point correlation function of rich clusters of galaxies: results from an extended APM cluster redshift survey', Oxford Univ. preprint, 1994, MNRAS, in press.

70. W. Zurek, *Ap. J.* **324**, 19 (1988).

71. M. Rees and J. Ostriker, *Mon. Not. R. astr. Soc.* **179**, 541 (1977).

72. Ya.B. Zel'dovich, J. Einasto and S. Shandarin, *Nature* **300**, 407 (1982);
J. Oort, *Ann. Rev. Astron. Astrophys.* **21**, 373 (1983);
R.B. Tully, *Ap. J.* **257**, 389 (1982);
S. Gregory, L. Thomson and W. Tifft, *Ap. J.* **243**, 411 (1980).

73. G. Chincarini and H. Rood, *Nature* **257**, 294 (1975);
J. Einasto, M. Joeveer and E. Saar, *Mon. Not. R. astron. Soc.* **193**, 353 (1980);
R. Giovanelli and M. Haynes, *Astron. J.* **87**, 1355 (1982);
D. Batuski and J. Burns, *Ap. J.* **299**, 5 (1985).

74. R. Kirshner, A. Oemler, P. Schechter and S. Shechtman, *Ap. J. (Lett.)* **248**, L57 (1981).

75. M. Joeveer, J. Einasto and E. Tago, *Mon. Not. R. astron. Soc.* **185**, 357 (1978);
L. da Costa et al., *Ap. J.* **327**, 544 (1988).

76. see e.g., G. Efstathiou, in 'Physics of the Early Universe,' proc. of the 1989 Scottish Univ. Summer School in Physics, ed. by J. Peacock, A. Heavens and A. Davies (SUSSP Publ., Edinburgh, 1990).

77. T. Padmanabhan, 'Structure Formation in the Universe' (Cambridge Univ. Press, Cambridge, 1993).

78. E. Harrison, *Phys. Rev.* **D1**, 2726 (1970);
Ya.B. Zel'dovich, *Mon. Not. R. astron. Soc.* **160**, 1p (1972).

79. E. Lifshitz, *Zh. Eksp. Teor. Fiz.* **16**, 587 (1946);
E. Lifshitz and I. Khalatnikov, *Adv. Phys.* **12**, 185 (1963).

80. W. Press and E. Vishniac, *Ap. J.* **239**, 1 (1980).

81. R. Brandenberger, H. Feldman, V. Mukhanov and T. Prokopec, 'Gauge Invariant Cosmological Perturbations: Theory and Applications,' publ. in "The Origin of Structure in the Universe," eds. E. Gunzig and P. Nardone (Kluwer, Dordrecht, 1993).

82. J. Bardeen, *Phys. Rev.* **D22**, 1882 (1980).

83. R. Brandenberger, R. Kahn and W. Press, *Phys. Rev.* **D28**, 1809 (1983).

84. H. Kodama and M. Sasaki, *Prog. Theor. Phys. Suppl.* No. 78, 1 (1984).

85. R. Durrer and N. Straumann, *Helvet. Phys. Acta* **61**, 1027 (1988).

86. D. Lyth and M. Mukherjee, *Phys. Rev.* **D38**, 485 (1988).

87. G.F.R. Ellis and M. Bruni, *Phys. Rev.* **D40**, 1804 (1989).

88. J. Stewart, *Class. Quantum Grav.* **7**, 1169 (1990).

89. J. Stewart and M. Walker, *Proc. R. Soc.* **A341**, 49 (1974).

90. J. Bardeen, P. Steinhardt and M. Turner, *Phys. Rev.* **D28**, 679 (1983).

91. D. Kirzhnits and A. Linde, *Pis'ma Zh. Eksp. Teor. Fiz.* **15**, 745 (1972);
D. Kirzhnits and A. Linde, *Zh. Eksp. Teor. Fiz.* **67**, 1263 (1974);
C. Bernard, *Phys. Rev.* **D9**, 3313 (1974);
L. Dolan and R. Jackiw, *Phys. Rev.* **D9**, 3320 (1974);
S. Weinberg, *Phys. Rev.* **D9**, 3357 (1974).

92. A. Linde, *Phys. Lett.* **129B**, 177 (1983).

93. G. Mazenko, W. Unruh and R. Wald, *Phys. Rev.* **D31**, 273 (1985).

94. A. Linde, *Phys. Lett.* **108B**, 389 (1982);
A. Albrecht and P. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).

95. J. Langer, *Physica* **73**, 61 (1974).

96. S. Coleman, *Phys. Rev.* **D15**, 2929 (1977);
C. Callan and S. Coleman, *Phys. Rev.* **D16**, 1762 (1977).

97. M. Voloshin, Yu. Kobzarev and L. Okun, *Sov. J. Nucl. Phys.* **20**, 644 (1975).

98. M. Stone, *Phys. Rev.* **D14**, 3568 (1976);
M. Stone, *Phys. Lett.* **67B**, 186 (1977).

99. P. Frampton, *Phys. Rev. Lett.*, **37**, 1380 (1976).

100. S. Coleman, in 'The Whys of Subnuclear Physics' (Erice 1977), ed by A. Zichichi (Plenum, New York, 1979).

101. A. Guth and S.-H. Tye, *Phys. Rev. Lett.* **44**, 631 (1980).

102. A. Guth and E. Weinberg, *Nucl. Phys.* **B212**, 321 (1983).

103. S. Hawking and I. Moss, *Phys. Lett.* **110B**, 35 (1982).

104. R. Matzner, in 'Proceedings of the Drexel Workshop on Numerical Relativity', ed by J. Centrella (Cambridge Univ. Press, Cambridge, 1986).

105. A. Albrecht, P. Steinhardt, M. Turner and F. Wilczek, *Phys. Rev. Lett.* **48**, 1437 (1982).

106. L. Abbott, E. Farhi and M. Wise, *Phys. Lett.* **117B**, 29 (1982).

107. J. Traschen and R. Brandenberger, *Phys. Rev.* **D42**, 2491 (1990).

108. L. Kofman, A. Linde and A. Starobinski, *Phys. Rev. Lett.* **73**, 3195 (1994); Y. Shtanov, J. Traschen and R. Brandenberger, *Phys. Rev.* **D51**, 5438 (1995).

109. L. Landau and E. Lifshitz, 'Mechanics' (Pergamon, Oxford, 1960); V. Arnold, 'Mathematical Methods of Classical Mechanics' (Springer, New York, 1978).

110. S. Coleman and E. Weinberg, *Phys. Rev.* **D7**, 1888 (1973).

111. M. Markov and V. Mukhanov, *Phys. Lett.* **104A**, 200 (1984); V. Belinsky, L. Grishchuk, I. Khalatnikov and Ya. Zel'dovich, *Phys. Lett.* **155B**, 232 (1985); L. Kofman, A. Linde and A. Starobinsky, *Phys. Lett.* **157B**, 36 (1985); T. Piran and R. Williams, *Phys. Lett.* **163B**, 331 (1985).

112. S.-Y. Pi, *Phys. Rev. Lett.* **52**, 1725 (1984); K. Freese, J. Frieman and A. Olinto, *Phys. Rev. Lett.* **65**, 3233 (1990).

113. D. Nanopoulos, K. Olive, M. Srednicki and K. Tamvakis, *Phys. Lett.* **123B**, 41 (1983);
J. Ellis, K. Enqvist, D. Nanopoulos, K. Olive and M. Srednicki, *Phys. Lett.* **152B**, 175 (1985);
R. Holman, P. Ramond and C. Ross, *Phys. Lett.* **137B**, 343 (1984).

114. A. Goncharov and A. Linde, *JETP* **59**, 930 (1984);
A. Goncharov and A. Linde, *Phys. Lett.* **139B**, 27 (1984);
A. Goncharov and A. Linde, *Class. Quant. Grav.* **1**, L75 (1984).

115. I. Antoniadis, J. Ellis, J. Hagelin and D. Nanopoulos, *Phys. Lett.* **205B**, 459 (1988);
I. Antoniadis, J. Ellis, J. Hagelin and D. Nanopoulos, *Phys. Lett.* **208B**, 209 (1988).

116. A. Linde, D. Linde and A. Mezhlumian, *Phys. Rev.* **D49**, 1783 (1994);
A. Linde, 'Lectures on Inflationary Cosmology', Stanford preprint SU-ITP-94-36, hep-th/9410082 (1994).

117. A. Starobinsky, in 'Current Trends in Field Theory, Quantum Gravity, and Strings', Lecture Notes in Physics, ed. by H. de Vega and N. Sanchez (Springer, Heidelberg, 1986).

118. B. Whitt, *Phys. Lett.* **145B**, 176 (1984).

119. M. Gasperini and G. Veneziano, *Astropart. Phys.* **1**, 317 (1993);
R. Brustein and G. Veneziano, *Phys. Lett.* **B329**, 429 (1994).

120. F. Adams, K. Freese and A. Guth, *Phys. Rev.* **D43**, 965 (1991).

121. H. Feldman and R. Brandenberger, *Phys. Lett.* **227B**, 359 (1989).

122. J. Kung and R. Brandenberger, *Phys. Rev.* **D42**, 1008 (1990).

123. D. Goldwirth and T. Piran, *Phys. Rev. Lett.* **64**, 2852 (1990);
D. Goldwirth and T. Piran, *Phys. Rep.* **214**, 223 (1992).

124. A. Albrecht and R. Brandenberger, *Phys. Rev.* **D31**, 1225 (1985).

125. G. Chibisov and V. Mukhanov, *JETP Lett.* **33**, 532 (1981);
V. Mukhanov and G. Chibisov, *Zh. Eksp. Teor. Fiz.* **83**, 475 (1982).

126. A. Lapedes, *J. Math. Phys.* **19**, 2289 (1978);
R. Brandenberger and R. Kahn, *Phys. Lett.* **119B**, 75 (1982).

127. A. Guth and S.-Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982);
S. Hawking, *Phys. Lett.* **115B**, 295 (1982);
A. Starobinsky, *Phys. Lett.* **117B**, 175 (1982);
R. Brandenberger and R. Kahn, *Phys. Rev.* **D28**, 2172 (1984);
J. Frieman and M. Turner, *Phys. Rev.* **D30**, 265 (1984);
V. Mukhanov, *JETP Lett.* **41**, 493 (1985).

128. W. Zurek, *Phys. Rev.* **D24**, 1516 (1982);
W. Zurek, *Phys. Rev.* **D26** 1862 (1982).

129. E. Joos and H. Zeh, *Z. Phys.* **B59**, 223 (1985);
H. Zeh, *Phys. Lett.* **116A**, 9 (1986);
C. Kiefer, *Class. Quantum Grav.* **4**, 1369 (1987);
T. Fukuyama and M. Morikawa, *Phys. Rev.* **D39**, 462 (1989);
J. Halliwell, *Phys. Rev.* **D39**, 2912 (1989);
T. Padmanabhan, *Phys. Rev.* **D39**, 2924 (1980);
W. Unruh and W. Zurek, *Phys. Rev.* **D40**, 1071 (1989);
E. Calzetta and F. Mazzitelli, *Phys. Rev.* **D42**, 4066 (1990);
S. Habib and R. Laflamme, *Phys. Rev.* **D42**, 4056 (1990);
H. Feldman and A. Kamenshchik, *Class. Quantum Grav.* **8**, L65 (1991).

130. M. Sakagami, *Prog. Theor. Phys.* **79**, 443 (1988);
R. Brandenberger, R. Laflamme and M. Mijic, *Mod. Phys. Lett.* **A5**, 2311 (1990).

131. J. Bardeen, unpublished (1984).

132. R. Brandenberger, *Nucl. Phys.* **B245**, 328 (1984).

133. N. Birrell and P. Davies, 'Quantum Fields in Curved Space' (Cambridge Univ. Press, Cambridge, 1982).

134. R. Brandenberger and C. Hill, *Phys. Lett.* **179B**, 30 (1986).

135. W. Fischler, B. Ratra and L. Susskind, *Nucl. Phys.* **B259**, 730 (1985).

136. D. Mermin, *Rev. Mod. Phys.* **51**, 591 (1979).

137. P. de Gennes, 'The Physics of Liquid Crystals' (Clarendon Press, Oxford, 1974);
I. Chuang, R. Durrer, N. Turok and B. Yurke, *Science* **251**, 1336 (1991);
M. Bowick, L. Chandar, E. Schiff and A. Srivastava, *Science* **263**, 943 (1994).

138. M. Salomaa and G. Volovik, *Rev. Mod. Phys.* **59**, 533 (1987).

139. A. Abrikosov, *JETP* **5**, 1174 (1957).

140. Ya.B. Zel'dovich, I. Kobzarev and L. Okun, *Zh. Eksp. Teor. Fiz.* **67**, 3 (1974).

141. Ya.B. Zel'dovich and M. Khlopov, *Phys. Lett.* **79B**, 239 (1978);
J. Preskill, *Phys. Rev. Lett.* **43**, 1365 (1979).

142. T.W.B. Kibble, *Acta Physica Polonica* **B13**, 723 (1982).

143. P. Langacker and S.-Y. Pi, *Phys. Rev. Lett.* **45**, 1 (1980).

144. T.W.B. Kibble and E. Weinberg, *Phys. Rev.* **D43**, 3188 (1991).

145. T. Vachaspati and A. Vilenkin, *Phys. Rev.* **D30**, 2036 (1984).

146. S. Rudaz and A. Srivastava, *Mod. Phys. Lett.* **A8**, 1443 (1993).

147. F. Liu, M. Mondello and N. Goldenfeld, *Phys. Rev. Lett.* **66**, 3071 (1991).

148. M. Hindmarsh, A.-C. Davis and R. Brandenberger, *Phys. Rev.* **D49**, 1944 (1994);
R. Brandenberger and A.-C. Davis, *Phys. Lett.* **B332**, 305 (1994).

149. T. Prokopec, A. Sornborger and R. Brandenberger, *Phys. Rev.* **D45**, 1971 (1992).

150. J. Borrill, E. Copeland and A. Liddle, *Phys. Lett.* **258B**, 310 (1991).

151. A. Sornborger, *Phys. Rev.* **D48**, 3517 (1993).

152. L. Perivolaropoulos, *Phys. Rev.* **D46**, 1858 (1992).

153. T. Prokopec, *Phys. Lett.* **262B**, 215 (1991);
R. Leese and T. Prokopec, *Phys. Rev.* **D44**, 3749 (1991).

154. M. Barriola and A. Vilenkin, *Phys. Rev. Lett.* **63**, 341 (1989).

155. S. Rhie and D. Bennett, *Phys. Rev. Lett.* **65**, 1709 (1990).

156. D. Foerster, *Nucl. Phys.* **B81**, 84 (1974).

157. N. Turok, in 'Proceedings of the 1987 CERN/ESO Winter School on Cosmology and Particle Physics' (World Scientific, Singapore, 1988).

158. T.W.B. Kibble and N. Turok, *Phys. Lett.* **116B**, 141 (1982).

159. R. Brandenberger, *Nucl. Phys.* **B293**, 812 (1987).

160. E.P.S. Shellard, *Nucl. Phys.* **B283**, 624 (1987).

161. R. Matzner, *Computers in Physics* **1**, 51 (1988);
K. Moriarty, E. Myers and C. Rebbi, *Phys. Lett.* **207B**, 411 (1988);
E.P.S. Shellard and P. Ruback, *Phys. Lett.* **209B**, 262 (1988).

162. P. Ruback, *Nucl. Phys.* **B296**, 669 (1988).

163. A. Albrecht and N. Turok, *Phys. Rev. Lett.* **54**, 1868 (1985).

164. D. Bennett and F. Bouchet, *Phys. Rev. Lett.* **60**, 257 (1988).

165. B. Allen and E.P.S. Shellard, *Phys. Rev. Lett.* **64**, 119 (1990).

166. A. Albrecht and N. Turok, *Phys. Rev.* **D40**, 973 (1989).

167. R. Brandenberger and J. Kung, in 'The Formation and Evolution of Cosmic Strings' eds. G. Gibbons, S. Hawking and T. Vachaspati (Cambridge Univ. Press, Cambridge, 1990).

168. see e.g., C. Misner, K. Thorne and J. Wheeler, 'Gravitation' (Freeman, San Francisco, 1973).

169. T. Vachaspati and A. Vilenkin, *Phys. Rev.* **D31**, 3052 (1985);
N. Turok, *Nucl. Phys.* **B242**, 520 (1984);
C. Burden, *Phys. Lett.* **164B**, 277 (1985).

170. B. Carter, *Phys. Rev.* **D41**, 3869 (1990).

171. E. Copeland, T.W.B. Kibble and D. Austin, *Phys. Rev.* **D45**, 1000 (1992).

172. N. Turok, *Nucl. Phys.* **B242**, 520 (1984).

173. J. Silk and V. Vilenkin, *Phys. Rev. Lett.* **53**, 1700 (1984).

174. A. Vilenkin, *Phys. Rev.* **D23**, 852 (1981);
J. Gott, *Ap. J.* **288**, 422 (1985);
W. Hiscock, *Phys. Rev.* **D31**, 3288 (1985);
B. Linet, *Gen. Rel. Grav.* **17**, 1109 (1985);
D. Garfinkle, *Phys. Rev.* **D32**, 1323 (1985);
R. Gregory, *Phys. Rev. Lett.* **59**, 740 (1987).

175. D. Vollick, *Phys. Rev.* **D45**, 1884 (1992);
T. Vachaspati and A. Vilenkin, *Phys. Rev. Lett.* **67**, 1057 (1991).

176. A. Albrecht and A. Stebbins, *Phys. Rev. Lett.* **68**, 2121 (1992).

177. A. Albrecht and A. Stebbins, *Phys. Rev. Lett.* **69**, 2615 (1992).

178. D. Spergel, N. Turok, W. Press and B. Ryden, *Phys. Rev.* **D43**, 1038 (1991).

179. A. Gooding, D. Spergel and N. Turok, *Ap. J. (Lett.)* **372**, L5 (1991);
C. Park, D. Spergel and N. Turok, *Ap. J. (Lett.)* **373**, L53 (1991).

180. R. Cen, J. Ostriker, D. Spergel and N. Turok, *Ap. J.* **383**, 1 (1991).

181. J. Gott, A. Melott and M. Dickinson, *Ap. J.* **306**, 341 (1986).

182. S. Ramsey, Senior thesis, Brown Univ. (1992);
D. Kaplan, Senior thesis, Brown Univ. (1993);
R. Brandenberger, D. Kaplan and S. Ramsey, 'Some statistics for measuring large-scale structure', Brown preprint BROWN-HET-922 (1993);
A. Aguirre, Senior thesis, Brown Univ. (1995).

183. D. Fixen, E. Cheng and D. Wilkinson, *Phys. Rev. Lett.* **50**, 620 (1983).

184. R. Sachs and A. Wolfe, *Ap. J.* **147**, 73 (1967).

185. V. Mukhanov and G. Chibisov, *Sov. Astron. Lett.* **10**, 890 (1984).

186. N. Kaiser and A. Stebbins, *Nature* **310**, 391 (1984).

187. R. Moessner, L. Perivolaropoulos and R. Brandenberger, *Ap. J.* **425**, 365 (1994).

188. N. Turok and D. Spergel, *Phys. Rev. Lett.* **64**, 2736 (1990).

189. R. Durrer, A. Howard and Z.-H. Zhou, *Phys. Rev.* **D49**, 681 (1994).

190. U.-L. Pen, D. Spergel and N. Turok, *Phys. Rev.* **D49**, 692 (1994).

191. E. Wright et al., *Ap. J. (Lett.)* **396**, L13 (1992).

192. S. White, C. Frenk, M. Davis and G. Efstathiou, *Ap. J.* **313**, 505 (1987).

193. D. Bennett and S. Rhie, *Ap. J. (Lett.)* **406**, L7 (1993).

194. D. Bennett, A. Stebbins and F. Bouchet, *Ap. J. (Lett.)* **399**, L5 (1992).

195. L. Perivolaropoulos, *Phys. Lett.* **298B**, 305 (1993).

196. G. Smoot et al., *Ap. J. (Lett.)* **396**, L1 (1992).

197. S. Meyer, E. Cheng and L. Page, *Ap. J. (Lett.)* **371**, L7 (1991);
K. Ganga, E. Cheng, S. Meyer and L. Page, *Ap. J. (Lett.)* **410**, L57 (1993).

198. S. Hancock et al., *Nature* **317**, 333 (1994).

199. E. Wollack et al., *Ap. J. (Lett.)* **419**, L49 (1993).

200. T. Gaier et al., *Ap. J. (Lett.)* **398**, L1 (1992);
J. Schuster et al., *Ap. J. (Lett.)* **412**, L47 (1993).

201. P. de Bernardis et al., *Ap. J. (Lett.)* **422**, L33 (1994).

202. M. Dragovan et al., Princeton preprint (1993).

203. P. Meinhold et al., *Ap. J. (Lett.)* **409**, L1 (1993).

204. J. Gunderson et al., *Ap. J. (Lett.)* **413**, L1 (1993).

205. E. Cheng et al., *Ap. J. (Lett.)* **422**, L37 (1994).

206. G. Tucker, G. Griffin, H. Nguyen and J. Peterson, *Ap. J. (Lett.)* **419**, L45 (1993).

207. A. Readhead et al., *Ap. J.* **346**, 566 (1989).

208. S. Myers, A. Readhead and A. Lawrence, *Ap. J.* **405**, 8 (1993).

209. W. Freedman et al., *Nature* **371**, 757 (1994);
M. Pierce et al., *Nature* **371**, 385 (1994).

210. A. Riess, W. Press and R. Kirshner, *Ap. J. (Lett.)* **438**, L17 (1995).

211. M. Jones et al., *Nature* **365**, 320 (1993).

212. R. Brandenberger, *Phys. Lett.* **B129**, 397 (1983).

213. N. Tsamis and R. Woodard, *Phys. Lett.* **B301**, 351 (1993).

214. W. Zurek, *Acta Phys. Pol.* **B24**, 1301 (1993).

215. T. Kibble and A. Vilenkin, "Density of strings formed at a second order cosmological phase transition", Imperial preprint IMPERIAL-TP-94/95-9A, hep-ph/9501207 (1995).

216. A. Aguirre and R. Brandenberger, "Accretion of hot dark matter onto slowly moving cosmic strings", Brown preprint BROWN-HET-995, astro-ph/9505031, *Int. J. Mod. Phys. D*, in press (1995).

217. R. Moessner and R. Brandenberger, "Formation of high redshift objects in a cosmic string model with hot dark matter", Brown preprint BROWN-HET-1001 (1995).

218. M. Trodden, V. Mukhanov and R. Brandenberger, *Phys. Lett.* **B316**, 483 (1993).

219. R. Moessner and M. Trodden, *Phys. Rev.* **D51**, 2801 (1995).

220. B. Altshuler, *Class. Quant. Grav.* **7**, 189 (1990).

221. M. Markov, *Pis'ma Zh. Eksp. Theor. Fiz.* **36**, 214 (1982);
M. Markov, *Pis'ma Zh. Eksp. Theor. Fiz.* **46**, 342 (1987);
V. Ginsburg, V. Mukhanov and V. Frolov, *Pis'ma Zh. Eksp. Theor. Fiz.* **94**, 3 (1988);
V. Frolov, M. Markov and V. Mukhanov, *Phys. Rev.* **D41**, 383 (1990).

222. R. Brandenberger, M. Mohazzab, V. Mukhanov, A. Sornborger and M. Trodden, in preparation (1995).

223. K. Kikkawa and M. Yamasaki, *Phys. Lett.* **B149**, 357 (1984);
N. Sakai and I. Senda, *Prog. Theor. Phys.* **75**, 692 (1986);
B. Sathiapalan, *Phys. Rev. Lett.* **58**, 1597 (1987);
P. Ginsparg and C. Vafa, *Nucl. Phys.* **B289**, 414 (1987).

224. R. Brandenberger and C. Vafa, *Nucl. Phys.* **B316**, 391 (1989).

225. R. Hagedorn, *Nuovo Cimento Suppl.* **3**, 147 (1965).

226. D. Mitchell and N. Turok, *Nucl. Phys.* **B294**, 1138 (1987).

227. N. Deo, S. Jain and C.-I. Tan, *Phys. Rev.* **D40**, 2626 (1989).

228. A. Tseytlin and C. Vafa, *Nucl. Phys.* **B372**, 443 (1992).

229. G. Veneziano, *Phys. Lett.* **B265**, 287 (1991).